Atomic Cluster Expansion

Muchang Bahng

April 2022

1 Molecular Dynamics

In this context we work with a system of N point particles in $\mathbb{R}^3$, with the state of each particle fully characterized by its position and momentum vectors. Let us denote the masses of the particles as m_i, which will be commonly represented as the $3N \times 3N$ matrix

$$\mathbf{M} = \text{diag}(m_1, \dots, m_N) \otimes I_3 = \begin{pmatrix} m_1 & & & & & & & & \\ & m_1 & & & & & & & \\ & & m_1 & & & & & & \\ & & & m_2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & m_{N-1} & & & \\ & & & & & & m_N & & \\ & & & & & & & m_N & \\ & & & & & & & & m_N \end{pmatrix},$$

the position vector of all particles as $\mathbf{q} = (\mathbf{q}_1, \dots, \mathbf{q}_N) \in \mathbb{R}^{3N}$, and the momentum vector of all particles as $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_N) \in \mathbb{R}^{3N}$. The collective kinetic energy of the system is

$$E_{\text{kin}}(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}$$

and hence the total energy/Hamiltonian of the particle system is

$$H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + E_{\text{kin}}(\mathbf{p})$$

Note that the potential energy depends only on the position vector $\mathbf{q}$, while the kinetic energy depends on the momentum $\mathbf{p}$.

Note that since $H : \Omega_{\mathbf{q}} \times \mathbb{R}^{3N} \longrightarrow \mathbb{R}$, its gradient vector can be represented as

$$\nabla H(\mathbf{q}, \mathbf{p}) = \begin{pmatrix} \nabla_{\mathbf{q}} H \\ \nabla_{\mathbf{p}} H \end{pmatrix} = \begin{pmatrix} \frac{\partial H}{\partial \mathbf{q}} \\ \frac{\partial H}{\partial \mathbf{p}} \end{pmatrix} = \begin{pmatrix} \partial H / \partial q_1 \\ \vdots \\ \partial H / \partial q_{3N} \\ \partial H / \partial p_1 \\ \vdots \\ \partial H / \partial p_{3N} \end{pmatrix}$$

But since $H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + E_{\text{kin}}(\mathbf{p})$ is separable and since

$$\begin{aligned} \nabla_{\mathbf{p}} E_{\text{kin}}(p) &= \nabla_{\mathbf{p}} \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} \\ &= \nabla_{\mathbf{p}} \frac{1}{2} (m_1^{-1} p_{11}^2 + m_1^{-1} p_{12}^2 + m_1^{-1} p_{13}^2 + m_2^{-1} p_{21}^2 + \dots + m_N^{-1} p_{N3}^2) \\ &= \left(m_1^{-1} p_{11},\ m_1^{-1} p_{12},\ m_1^{-1} p_{13},\ m_2^{-1} p_{21}, \dots, m_N^{-1} p_{N3}\right)^T \\ &= \mathbf{M}^{-1} \mathbf{p} \end{aligned}$$

we have

$$\nabla H(\mathbf{q},\mathbf{p}) = \begin{pmatrix} \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \nabla_{\mathbf{p}} E_{\text{kin}}(\mathbf{p}) \end{pmatrix} = \begin{pmatrix} \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \mathbf{M}^{-1}\mathbf{p} \end{pmatrix}$$

and therefore, the equations of motions can be rewritten as

$$\begin{cases} \dot{\boldsymbol{q}} &= \mathbf{M}^{-1}\mathbf{p} \\ \dot{\boldsymbol{p}} &= -\nabla_{\mathbf{q}} U(\mathbf{q}) \end{cases} \implies \begin{pmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{3N} \\ -\mathbf{I}_{3N} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \nabla_{\mathbf{q}} U(\mathbf{q}) \\ \mathbf{M}^{-1}\mathbf{p} \end{pmatrix} = \mathbf{J}\nabla H(\mathbf{q},\mathbf{p})$$

1.1 Hamiltonian Flow Maps

The particles are assumed to obey the laws of classical mechanics, and so the time evolution of the particle system is described by Newton's equations of motion, i.e.

$$\begin{aligned} \dot{\boldsymbol{q}} &= \mathbf{M}^{-1}\mathbf{p} \\ \dot{\boldsymbol{p}} &= \mathbf{F}(q) = -\nabla_{\mathbf{q}} U(\mathbf{q}) \end{aligned}$$

With $\mathbf{q} \in \Omega_{\mathbf{q}} \subset \mathbb{R}^{3N}$ and $\mathbf{p} \in \mathbb{R}^{3N}$, the configuration space is just

$$\Omega_{\mathbf{q}} \times \mathbb{R}^{3N}$$

Now, let $\Phi_t : \Omega_{\mathbf{q}} \times \mathbb{R}^{3N} \longrightarrow \Omega_{\mathbf{q}} \times \mathbb{R}^{3N}$ denote the flow map associated with Newton's equations, defined such that

$$\Phi_t\big(\mathbf{q}(0),\mathbf{p}(0)\big) \coloneqq \big(\mathbf{q}(t),\mathbf{p}(t)\big)$$

for any initialization $\big(\mathbf{q}(0),\mathbf{p}(0)\big) \in \Omega_{\mathbf{q}} \times \mathbb{R}^{3N}$. In other words,

$$\Phi_t\big(\mathbf{q}(0),\mathbf{p}(0)\big) : \mathbb{R} \longrightarrow \Omega_{\mathbf{q}} \times \mathbb{R}^{3N}$$

defines a full evolution path of the system given the original point. It has a few properties.

1. The collection of flow maps form an algebraic group in the sense that

   $$\Phi_t \cdot \Phi_s = \Phi_{t+s}$$

   with the identity element $\Phi_0 = \mathrm{Id}$ (the path map that doesn't go anywhere), and well-defined inverse

   $$\Phi_t^{-1} = \Phi_{-t}$$

   In this case this group is isomorphic to $(\mathbb{R},+)$.

2. Symmetry holds in the sense that

   $$S \circ \Phi_t \circ S = \Phi_{-t}$$

   where the function $S : (\mathbf{q},\mathbf{p}) \mapsto (\mathbf{q},-\mathbf{p})$ flips the momentum.

3. Total energy is conserved under Φ_t.

   $$H\big(\mathbf{q}(t),\mathbf{p}(t)\big) = H\big(\mathbf{q}(0),\mathbf{p}(0)\big)$$

4. In the absence of an external force, the total momentum is conserved under Φ_t.

1.1.1 The Symplectic Property

The final property is less obvious. A fundamental property of solutions of Hamiltonian differential equations is that the collection $(\Phi_t)_{t\in\mathbb{R}}$ of associated flow maps has a symplectic group structure, which means that the symplectic 2-form is preserved under the action of each group element.

1. A **1-form** α defined on $\mathbb{R}^{6N}$ is a family of linear mappings such that for every $\mathbf{x} \in \mathbb{R}^{6N}$, $\alpha(\mathbf{x})$ is a linear map from $\mathbb{R}^{6N}$ to $\mathbb{R}$. That is, given a linear map $\mathbf{a} : \mathbb{R}^{6N} \longrightarrow \mathbb{R}^{6N}$, we may define a one-form associated to this vector field $\mathbf{a} \mapsto \alpha$ by

$$\alpha(\mathbf{x})(\boldsymbol{\xi}) = \mathbf{a}(\mathbf{x})^T\boldsymbol{\xi}$$

2. The **differential** of a function $g : \mathbb{R}^{6N} \longrightarrow \mathbb{R}$, denoted $\mathrm{d}g$, is a family of linear mappings from vectors $\boldsymbol{\xi} \in \mathbb{R}^{6N}$ into the reals defined by

$$\mathrm{d}g(\mathbf{q},\mathbf{p})(\boldsymbol{\xi}) = \nabla g(\mathbf{q},\mathbf{p})^T\boldsymbol{\xi}$$

   Therefore, we can see that the differential is an example of a 1-form.

3. The wedge product of 1-forms α, β is a **2-form**, which can be viewed as a quadratic form, i.e. a scalar-valued function of two vectors which is linear in each argument. It is written $\alpha \wedge \beta$ and is defined, for vectors $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^{6N}$ by

$$(\alpha \wedge \beta)(\boldsymbol{\xi}, \boldsymbol{\eta}) \coloneqq \alpha(\boldsymbol{\xi})\beta(\boldsymbol{\eta}) - \alpha(\boldsymbol{\eta})\beta(\boldsymbol{\xi})$$

Now, let $q_i, p_j : \mathbb{R}^{6N} \longrightarrow \mathbb{R}$ be the component functions mapping $(\mathbf{q},\mathbf{p}) \mapsto q_i, p_j$, respectively, where $1 \le i, j \le 3N$. Then, $\mathrm{d}q_i, \mathrm{d}p_i$ are examples of differential 1-forms. The wedge product of the coordinate differentials $\mathrm{d}q_i, \mathrm{d}p_i$ can be written

$$(\mathrm{d}q_i \wedge \mathrm{d}p_i)(\boldsymbol{\xi}, \boldsymbol{\eta}) = \xi_i\eta_{i+3N} - \xi_{i+3N}\eta_i = \boldsymbol{\xi}^T\mathbf{J}^{(i)}\boldsymbol{\eta}$$

where $\mathbf{J}$ is the matrix which has zeros everywhere except for $(\mathbf{J}^{(i)})_{i,3N+i} = 1, (\mathbf{J}^{(i)})_{i+3N,i} = -1$. Summing these terms results in the symplectic 2-form, denoted ψ_S:

$$\psi_S \coloneqq \sum_{i=1}^{3N} \mathrm{d}q_i \wedge \mathrm{d}p_i(\boldsymbol{\xi}, \boldsymbol{\eta}) = \boldsymbol{\xi}^T\left(\sum_{i=1}^{3N}\mathbf{J}^{(i)}\right)\boldsymbol{\eta} = \boldsymbol{\xi}^T\mathbf{J}\boldsymbol{\eta}$$

That is, since $\Phi_t : \mathbb{R}^{6N} \longrightarrow \mathbb{R}^{6N}$, then its Jacobian $\nabla\Phi_t$ is a $6N \times 6N$ matrix, and the condition above implies that

$$\nabla\Phi_t^T\mathbf{J}\nabla\Phi_t = \mathbf{J} \text{ for all } t \in \mathbb{R}$$

Denoting the Jacobian $\nabla\Phi_t$ as Φ_t', we can take the determinant of both sides to find that

$$\det\left(\nabla\Phi_t^T\mathbf{J}\nabla\Phi_t\right) = \det(\mathbf{J}) \implies \det(\nabla\Phi_t^T)\det(\mathbf{J})\det(\nabla\Phi_t) = \det(\mathbf{J})$$

and so $\det{\Phi_t'}^2 = 1$. In the case of a flow map, we know that for $t \to 0$, the flow map Φ_t would essentially reduce to the identity map Id, and so

$$\lim_{t\to 0}\Phi_t' = 1 \implies \det\Phi_t' = +1$$

due to the determinant being a continuous function of t. Therefore a consequence of this is that Φ_t is volume preserving.

2 Numerical Integrators

By letting $\mathbf{z}$ represent the collection of all positions and momenta, we can represent the differential equation as

$$\begin{pmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{pmatrix} = \mathbf{J}\nabla H(\mathbf{q},\mathbf{p}) \implies \mathbf{z} = \mathbf{f}(\mathbf{z}), \text{ with } \mathbf{f}(\mathbf{z}) = \mathbf{J}\nabla H(\mathbf{z})$$

Solving this initial value problem with $\mathbf{z}(0) = \boldsymbol{\zeta}$ relies on the idea of a discretization with a finite stepsize h, and an iterative procedure $\hat{\Phi}$ that computes, starting from $\mathbf{z}_0 = \boldsymbol{\zeta}$, a sequence $\mathbf{z}_1, \mathbf{z}_2, \ldots$, where $\mathbf{z}_n \approx \mathbf{z}(nh)$. The simplest scheme is simply Euler's method which advances the solution from timestep to timestep by the formula

$$\mathbf{z}_{n+1} = \mathbf{z}_n + h\mathbf{f}(\mathbf{z}_n)$$

which is based on the observation that $\mathbf{z}(t+h) \approx \mathbf{z}(t) + h\,\dot{\boldsymbol{z}}(t)$, i.e. the beginning of a Taylor series expansion in powers of h, and the further observation that the solution satisfies the differential equation, hence $\dot{\boldsymbol{z}}(t)$ may be replaced by $\mathbf{f}(\mathbf{z}(t))$. We will focus on a particular class of schemes called **generalized one-step methods**. Letting the solution of the initial value problem $\dot{\boldsymbol{z}} = \mathbf{f}(\mathbf{z}), \mathbf{z}(0) = \boldsymbol{\zeta}$ be written as $\mathbf{z}(t, \boldsymbol{\zeta})$ (with $\mathbf{z}(0, \boldsymbol{\zeta}) = \boldsymbol{\zeta}$), the flow map Φ defined above satisfies

$$\Phi_t(\boldsymbol{\zeta}) = \mathbf{z}(t, \zeta)$$

Let us talk about the convergence of these schemes. The **order of accuracy** is the exponent in the power law by which the error in the method is related to the stepsize. For example, when we say that a scheme is of third order, we mean that the global error (on a fixed finite-time interval) can be bounded by Kh^3, where h is a sufficiently small timestep and K is a number which depends on the length of the time interval and the features of the problem, but which is independent of h. It is known that the Euler method is a first order method. That is, let the length of the time interval be τ, with the step size h and number of steps ν. Then, $\nu h = \tau$, and defining the error at step n to be $e_n \coloneqq ||\mathbf{z}_n - \mathbf{z}(t_n)||$, where $t_n = nh$, the maximum error of the Euler method at step ν satisfies

$$\bar{e} \coloneqq \max_{0<n\leq\nu} e_n \leq C(\tau)h$$

Using the order notation, we have $\bar{e} = \mathcal{O}(h)$. In simulations, it is apparent that the errors are larger at the end of the interval than at earlier times. We know that molecular dynamics trajectories need to be very long compared to the time-step used in order for them to be useful, so how the error grows in long simulations is quite important.

2.1 Higher Order Methods

One approach to higher accuracy is to to decrease the step size, but a more efficient way is to use a higher order method, which must satisfy a global error estimate (for finite time intervals) of the form

$$\bar{e} \approx C(\tau)\, h^r$$

For example, the Taylor series expansion of the solution may be written

$$\mathbf{z}(t+h) = \mathbf{z}(t) + h\dot{\boldsymbol{z}}(t) + \frac{1}{2}h^2\ddot{\boldsymbol{z}}(t) + \ldots$$

and while truncating after the first term leads to Euler's method, truncating after the second order term leads to

$$\mathbf{z}_{n+1} = \mathbf{z}_n + h\dot{\boldsymbol{z}}_n + \frac{1}{2}h^2\ddot{\boldsymbol{z}}_n$$

In this formula, the second derivative is obtained by differentiating the differential equation itself:

$$\ddot{\boldsymbol{z}}(t) = \frac{d}{dt}\dot{\boldsymbol{z}}(t) = \frac{d}{dt}\mathbf{f}'\big(\mathbf{z}(t)\big) = \mathbf{f}'\big(\mathbf{z}(t)\big)\,\mathbf{f}\big(\mathbf{z}(t)\big)$$

So we can write the Taylor series method as below, which describes the flow map approximation:

$$\hat{\Phi}(\mathbf{z}_n) = \mathbf{z}_{n+1} = \mathbf{z}_n + h\,\mathbf{f}(\mathbf{z}_n) + \frac{1}{2}h^2\mathbf{f}'(\mathbf{z}_n)\,\mathbf{f}(\mathbf{z}_n)$$

Note that by the notation $\mathbf{f}'(\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^{6N}$ and $\mathbf{f} : \mathbb{R}^{6N} \longrightarrow \mathbb{R}^{6N}$ is the $6N \times 6N$ Jacobian matrix with ij-component being

$$[\mathbf{f}'(\mathbf{z})]_{ij} \coloneqq \frac{\partial f_i}{\partial z_j}$$

An alternative notation for $\mathbf{f}'$ is $\partial\mathbf{f}/\partial\mathbf{z}$. This higher order is an improvement, but they do not necessarily resolve the more important issue of unlimited growth of perturbations from the energy surface. An example of a second order method is the velocity Verlet method, described as

$$\begin{aligned}
\mathbf{p}_{k+1/2} &= \mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k) \\
\mathbf{q}_{k+1} &= \mathbf{q}_k + h\,\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\
\mathbf{p}_{k+1} &= \mathbf{p}_{k+1/2} - \frac{h}{2}\nabla U(\mathbf{q}_{k+1})
\end{aligned}$$

While this algorithm may not look like a second order, we can see that with simple substitution, we have a second order evaluation of $\mathbf{q}_{k+1}$ followed by an evaluation of $\mathbf{p}_{k+1}$.

$$\begin{aligned}
\mathbf{q}_{k+1} &= \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\
&= \mathbf{q}_k + h\mathbf{M}^{-1}\left(\mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k)\right) \\
&= \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_k - \frac{1}{2}h^2\nabla U(\mathbf{q}_k) \\
\mathbf{p}_{k+1} &= \mathbf{p}_{k+1/2} - \frac{h}{2}\nabla U(\mathbf{q}_{k+1}) \\
&= \left(\mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k)\right) - \frac{h}{2}\nabla U(\mathbf{q}_{k+1}) \\
&= \mathbf{p}_k - \frac{h}{2}\big[\nabla U(\mathbf{q}_k) + \nabla U(\mathbf{q}_{k+1})\big]
\end{aligned}$$

2.2 Convergence and the Order of Accuracy

A typical integrator computes successive steps from the formulas

$$\mathbf{z}_{n+1} = \hat{\Phi}_h(\mathbf{z}_n), \qquad \mathbf{z}_0 = \boldsymbol{\zeta}$$

Assume that $\hat{\Phi}_h$ is a smooth map for all $h > 0$. The exact solution $\mathbf{z}$ satisfies

$$\mathbf{z}(t_{n+1}) = \Phi_h(\mathbf{z}(t_n))$$

since Φ_h simply takes point $\mathbf{z}(t_n)$ and flows it for time period h. Therefore to each $h > 0$ we can associate a finite set of space points $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_\nu$, which represents the numerical solutions at $t_0 = 0, t_1 = h, t_2 = 2h, \ldots, t_\nu = \nu h = \tau$. Taking the difference of the numerical and exact solutions, we have

$$\mathbf{z}_{n+1} - \mathbf{z}(t_{n+1}) = \hat{\Phi}(\mathbf{z}_n) - \Phi_h\big(\mathbf{z}(t_n)\big) \tag{2.1}$$

We first assume that $\hat{\Phi}_h$ is an $\mathcal{O}(h^{p+1})$ approximation of Φ in the sense that there is a constant $K \geq 0$ and a constant $\Delta > 0$ such that, for $t \in [0, \tau]$, we have

$$||\Phi_t\big(\mathbf{z}(t)\big) - \hat{\Phi}_h\big(\mathbf{z}(t)\big)|| \leq Kh^{p+1}, \qquad h < \Delta$$

This assumption is usually verified by expanding the numerical and exact solutions in powers of h, using Taylor series expansions. To tackle the question of growth of error, we make an additional assumption on $\hat{\Phi}_h$, namely that is satisfies a **Lipshitz condition** of the form

$$||\hat{\Phi}_h(\mathbf{u}) - \hat{\Phi}_h(\mathbf{w})|| \leq (1 + hL)||\mathbf{u} - \mathbf{w}||$$

for all $\mathbf{u}, \mathbf{w} \in D$, where D is some subdomain of $\mathbb{R}^{6N}$ that contains the exact solution for $[0, \tau]$, and $h \leq \Delta$. This stability assumption, in words, says that the method does not increase the separation between two nearby trajectories by more than a factor of the form $1 + hL$. Therefore, from (2.1), we can write

$$\mathbf{z}_{n+1} - \mathbf{z}(t_{n+1}) = \hat{\Phi}(\mathbf{z}_n) - \hat{\Phi}_h\big(\mathbf{z}(t_n)\big) + \hat{\Phi}_h\big(\mathbf{z}(t_n)\big) - \Phi_h\big(\mathbf{z}(t_n)\big)$$

and take norms with the triangle inequality to get the following, where $\varepsilon_n = ||\mathbf{z}_n - \mathbf{z}(t_n)||$.

$$\varepsilon_{n+1} \leq (1 + Lh)\varepsilon_n + \bar{K}h^{p+1}$$

and from this, we can calculate the bound

$$\varepsilon_n \leq \frac{\bar{K}}{L} e^{Lnh} h^p$$

2.2.1 Verlet Method as 2nd Order Scheme

Recall that the Verlet method can be written in the form of a map as

$$\begin{aligned}
\mathbf{q}_{k+1} &= \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_k - \frac{1}{2}h^2\nabla U(\mathbf{q}_k) \\
\mathbf{p}_{k+1} = &= \mathbf{p}_k - \frac{h}{2}\big[\nabla_{\mathbf{q}} U(\mathbf{q}_k) + \nabla_{\mathbf{q}} U(\mathbf{q}_{k+1})\big]
\end{aligned}$$

Setting, $\mathbf{M} = I$, $\mathbf{F} = -\nabla_{\mathbf{q}} U$ and expanding the $\mathbf{q}_{k+1}$ in the inner term, we get

$$\begin{aligned}
\mathbf{q}_{k+1} &= \mathbf{q}_k + h\mathbf{p}_k + \frac{1}{2}h^2\mathbf{F}(\mathbf{q}_k) \\
\mathbf{p}_{k+1} &= \mathbf{p}_k + \frac{h}{2}\left[\mathbf{F}(\mathbf{q}_k) + \mathbf{F}\Big(\mathbf{q}_k + h\mathbf{p}_k + \frac{1}{2}h^2\mathbf{F}(\mathbf{q}_k)\Big)\right]
\end{aligned}$$

The first equation is already a polynomial, i.e. it is in the form of a series expansion in powers of h where the coefficients are functions of the starting point $(\mathbf{q}_k, \mathbf{p}_k)$. The second equation may be written as a series expansion in powers of h as well.

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \frac{h}{2}\mathbf{F}(\mathbf{q}_k) + \frac{h}{2}\left[\mathbf{F}(\mathbf{q}_k) + h\mathbf{F}'(\mathbf{q}_k)\big(\mathbf{p}_k + \frac{h}{2}\mathbf{F}(\mathbf{q}_k)\big) + \frac{h^2}{2}\mathbf{F}''(\mathbf{q}_k)\big(\mathbf{p}_k + \frac{h}{2}\mathbf{F}(\mathbf{q}_k)\big)^2 + \dots\right]$$

which we will neglect terms involving 4th and higher powers of h. Combining terms of like powers of h, we have

$$\mathbf{p}_{k+1} = \mathbf{p}_k + h\mathbf{F}(\mathbf{q}_k) + \frac{h^2}{2}\mathbf{p}_k\mathbf{F}'(\mathbf{q}_k) + \frac{h^3}{4}\big[\mathbf{F}'(\mathbf{q}_k)\mathbf{F}(\mathbf{q}_k) + \mathbf{p}_k^2\mathbf{F}''(\mathbf{q}_k)\big] + \mathcal{O}(h^4)$$

We compare this against the Taylor expansion of the exact solution $\mathbf{z}(t) \coloneqq \big(\mathbf{q}(t), \mathbf{p}(t)\big)$. We evaluate

$$\begin{aligned}
\mathbf{q}(t+h) &= \mathbf{q}(t) + h\mathbf{q}'(t) + \frac{h^2}{2}\mathbf{q}''(t) + \frac{h^3}{6}\mathbf{q}'''(t) + \mathcal{O}(h^4) \\
&= \mathbf{q}(t) + h\mathbf{p}(t) + \frac{h^2}{2}\mathbf{F}(\mathbf{q}(t)) + \frac{h^3}{6}\mathbf{F}'(\mathbf{q}(t))\,\mathbf{p}(t) + \mathcal{O}(h^4)
\end{aligned}$$

where the calculations followed from the fact that $\mathbf{q}' = \mathbf{p}$, which means that $\mathbf{q}'' = \mathbf{p}' = \mathbf{F}(\mathbf{q})$, which means that $\mathbf{q}''' = \frac{d}{dt}\mathbf{F}(\mathbf{q}) = \mathbf{F}'(\mathbf{q})\,\mathbf{q}' = \mathbf{F}'(\mathbf{q})\,\mathbf{p}$. Then, we have

$$\begin{aligned}\mathbf{p}(t+h) &= \mathbf{p}(t) + h\mathbf{p}'(t) + \frac{h^2}{2}\mathbf{p}''(t) + \frac{h^3}{6}\mathbf{p}'''(t) + \mathcal{O}(h^4)\\ &= \mathbf{p}(t) + h\mathbf{F}(\mathbf{q}(t)) + \frac{h^2}{2}\mathbf{p}(t)\,\mathbf{F}'(\mathbf{q}(t)) + \frac{h^3}{6}\big[\mathbf{p}(t)^2\,\mathbf{F}''(\mathbf{q}(t)) + \mathbf{F}'(\mathbf{q}(t))\,\mathbf{F}(\mathbf{q}(t))\big] + \mathcal{O}(h^4)\end{aligned}$$

which follows from the fact that $\mathbf{p}''' = (\mathbf{p}\,\mathbf{F}')' = \mathbf{p}^2\,\mathbf{F}'' + \mathbf{F}'\mathbf{F}$.
Now, let's compare them. Let us have initial point $\mathbf{z}_k = \mathbf{z}(t) = \big(\mathbf{q}(t), \mathbf{p}(t)\big) = \big(\mathbf{q}_k, \mathbf{p}_k\big)$ at time t. The actual flow and the integrator takes in $\mathbf{z}(t)$ and $\mathbf{z}_k$, respectively, but they are the same initial point, so we will label them with $\mathbf{z} = (\mathbf{q}, \mathbf{p})$. We can use the flow map $\Phi_h\big(\mathbf{z}(t)\big)$ to evaluate the exact position $\mathbf{z}(t+h)$ after time h. That is, $\Phi_h\big(\mathbf{z}(t)\big) \coloneqq \mathbf{z}\big(h, \mathbf{z}(t)\big) = \mathbf{z}(t+h)$. The Taylor expansion of this flow map is

$$\Phi_h\big(\mathbf{z}(t)\big) = \mathbf{z}(t+h) = \begin{cases}\mathbf{q}(t+h) = \mathbf{q} + h\mathbf{p} + \frac{h^2}{2}\mathbf{F} + \frac{h^3}{6}\mathbf{F}'\,\mathbf{p} + \mathcal{O}(h^4)\\ \mathbf{p}(t+h) = \mathbf{p} + h\mathbf{F} + \frac{h^2}{2}\mathbf{p}\,\mathbf{F}' + \frac{h^3}{6}\big[\mathbf{p}^2\,\mathbf{F}'' + \mathbf{F}'\,\mathbf{F}\big] + \mathcal{O}(h^4)\end{cases}$$

The numerical integrator would calculate something slightly different. That is, given the initial point $\mathbf{z}(t) = \mathbf{z}_k$, $\hat{\Phi}_h\big(\mathbf{z}(t)\big) = \mathbf{z}_{k+1}$ is the numerical approximation after time h. The Taylor expansion of this integrator is

$$\hat{\Phi}_h\big(\mathbf{z}(t)\big) = \mathbf{z}_{k+1} = \begin{cases}\mathbf{q}_{k+1} = \mathbf{q} + h\mathbf{p} + \frac{h^2}{2}\mathbf{F}\\ \mathbf{p}_{k+1} = \mathbf{p} + h\mathbf{F} + \frac{h^2}{2}\mathbf{p}\mathbf{F}' + \frac{h^3}{4}\big[\mathbf{F}'\mathbf{F} + \mathbf{p}^2\mathbf{F}''\big] + \mathcal{O}(h^4)\end{cases}$$

We should get $\mathbf{z}_{k+1} \approx \mathbf{z}(t+h)$, by looking at the differences, we find that these differ in the third (and higher) order terms.

$$\begin{aligned}\mathbf{q}_{k+1} - \mathbf{q}(t+h) &= -\frac{h^3}{6}\mathbf{F}'\mathbf{p} + \mathcal{O}(h^4)\\ \mathbf{p}_{k+1} - \mathbf{p}(t+h) &= \frac{h^3}{12}\big[\mathbf{p}^2\mathbf{F}'' + \mathbf{F}'\mathbf{F}\big] + \mathcal{O}(h^4)\end{aligned}$$

We can, in summary, state that the *local* error is third order

$$||\hat{\Phi}_h(\mathbf{z}) - \Phi_h(\mathbf{z})|| = \kappa(\mathbf{z})h^3 + \mathcal{O}(h^4)$$

where $\kappa(\mathbf{z}) \coloneqq \kappa(\mathbf{q}, \mathbf{p})$ is a function of the position and momentum. We may then define the maximum local error as

$$\bar{K} \coloneqq \max_{t\in[0,\tau]} \kappa\big(\mathbf{z}(t)\big)$$

and summing this all up leads to the total error being bounded by a constant multiple of h^2, achieving consistency of order 2.

3 Symplectic Integration

We wish to solve for Φ_t numerically by constructing an approximation with acceptable error.

$$(\hat{\mathbf{q}}_{n+1}, \hat{\mathbf{p}}_{n+1}) = \hat{\Phi}_h(\hat{\mathbf{q}}_n, \hat{\mathbf{p}}_n)$$

with $(\hat{\mathbf{q}_0}, \hat{\mathbf{p}_0}) = \big(\mathbf{q}(0), \mathbf{p}(0)\big)$. There is the obvious error stemming from the choice of a large h, but what is more important is that the geometric structure of the manifold $\big(\mathbf{q}(t), \mathbf{p}(t)\big)_{t>0}$ corresponding to the trajectory of the exact solution is replicated by the discrete approximation $\big(\hat{\mathbf{q}}_n, \hat{\mathbf{p}}_n\big)_{n\in\mathbb{N}}$. The best way to do this is to construct such a structure preserving integration scheme by designing the integration map $\hat{\Phi}_h$ in such a way that the symplectic 2-form is preserved. That is, construct a **symplectic integration scheme** $\hat{\Phi}_h$ such that

$$(\hat{\Phi}_h')^T\mathbf{J}\;\hat{\Phi}_h' = \mathbf{J}$$

3.1 Introduction to Stromer-Verlet and Velocity-Verlet Schemes

One of the most commonly used symplectic numerical integrators is the **Stormer-Verlet method**, which in algorithmic form reads

$$\begin{aligned}\mathbf{q}_{k+1/2} &= \mathbf{q}_k + \frac{h}{2}\mathbf{M}^{-1}\mathbf{p}_k \\ \mathbf{p}_{k+1} &= \mathbf{p}_k - h\,\nabla U(\mathbf{q}_{k+1/2}) \\ \mathbf{q}_{k+1} &= \mathbf{q}_{k+1/2} + \frac{h}{2}\mathbf{M}^{-1}\mathbf{q}_{k+1}\end{aligned}$$

We can see that this algorithm updates $\mathbf{q}_k \mapsto \mathbf{q}_{k+1/2}$ with the given $\mathbf{p}_k$ over half-time step, and then updates the force field vector with the new position vector $-\nabla U(\mathbf{q}_{k+1/2})$. This new force is used to update the momentum $\mathbf{p}_k \mapsto \mathbf{p}_{k+1}$. Finally, the position is updated with the new momentum: $\mathbf{q}_{k+1/2} \mapsto \mathbf{q}_{k+1}$. A closely related, alternative form is the **Velocity-Verlet method**, which updates the momentum first, then position, and finally momentum.

$$\begin{aligned}\mathbf{p}_{k+1/2} &= \mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k) \\ \mathbf{q}_{k+1} &= \mathbf{q}_k + h\,\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\ \mathbf{p}_{k+1} &= \mathbf{p}_{k+1/2} - \frac{h}{2}\nabla U(\mathbf{q}_{k+1})\end{aligned}$$

3.2 Symplectic Euler Integration

The standard Euler integration scheme has many shortfalls, such as error growth and stability issues. Its algorithmic form reads

$$\begin{aligned}\mathbf{q}_{k+1} &= \mathbf{q}_k + h\,\mathbf{M}^{-1}\mathbf{p}_k \\ \mathbf{p}_{k+1} &= \mathbf{p}_k - h\,\nabla U(\mathbf{q}_k)\end{aligned}$$

Therefore, modified version of this scheme, called the **symplectic Euler integration scheme**, is used, which reads

$$\begin{aligned}\mathbf{p}_{k+1} &= \mathbf{p}_k - h\,\nabla U(\mathbf{q}_k) \\ \mathbf{q}_{k+1} &= \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_{k+1}\end{aligned}$$

which has the slight modification that to advance the timestep, we use the first equation to compute $\mathbf{p}_{k+1}$ and then insert this in the second.

3.3 The Adjoint Method

Given any numerical integrator $\hat{\Phi}_h$, consider the map

$$(\hat{\Phi}_h)^\dagger := \hat{\Phi}_{-h}^{-1}$$

This is popularly referred to as the **adjoint** of $\hat{\Phi}_h$. For the flow map Φ_t, we know that the inverse map is precisely Φ_{-t}, so $\Phi_t^\dagger = \Phi_t$. That is, the flow map is in the normal sense self-adjoint, i.e. symmetric. However, such a property does not hold in the general case. Consider Euler's method $\hat{\Phi}_h$ in fully general form (Note that for the Euler method above, we have let $\mathbf{z} = (\mathbf{q}, \mathbf{p})^T$). Then, we have

$$\begin{aligned}&\hat{\Phi}_h : \mathbf{z}_k \mapsto \mathbf{z}_{k+1} \text{ such that } \mathbf{z}_{k+1} = \mathbf{z}_k + h\,\mathbf{f}(\mathbf{z}_k) \\ &\hat{\Phi}_h^{-1} : \mathbf{z}_k \mapsto \mathbf{z}_{k+1} \text{ such that } \mathbf{z}_k = \mathbf{z}_{k+1} + h\,\mathbf{f}(\mathbf{z}_{k+1}) \\ &\hat{\Phi}_{-h}^{-1} : \mathbf{z}_k \mapsto \mathbf{z}_{k+1} \text{ such that } \mathbf{z}_k = \mathbf{z}_{k+1} - h\,\mathbf{f}(\mathbf{z}_{k+1}) \iff \mathbf{z}_{k+1} = \mathbf{z}_k + h\,\mathbf{f}(\mathbf{z}_{k+1})\end{aligned}$$

and clearly, $\hat{\Phi}_{-h}^{-1}$ defines $\mathbf{z}_{k+1}$ implicitly. To construct the adjoint method of the symplectic Euler scheme, we see that that $\hat{\Phi}_h^{-1}$ maps $(\mathbf{q}_k, \mathbf{p}_k) \mapsto (\mathbf{q}_{k+1}, \mathbf{p}_{k+1})$ such that

$$\begin{aligned}\mathbf{p}_k &= \mathbf{p}_{k+1} - h\,\nabla U(\mathbf{q}_{k+1})\\ \mathbf{q}_k &= \mathbf{q}_{k+1} + h\,\mathbf{M}^{-1}\mathbf{p}_k\end{aligned}$$

and therefore $\hat{\Phi}_{-h}^{-1}$ maps $(\mathbf{q}_k, \mathbf{p}_k) \mapsto (\mathbf{q}_{k+1}, \mathbf{p}_{k+1})$ such that

$$\begin{aligned}\mathbf{q}_{k+1} &= \mathbf{q}_k + h\,\mathbf{M}^{-1}\mathbf{p}_k\\ \mathbf{p}_{k+1} &= \mathbf{p}_k - h\,\nabla U(\mathbf{q}_{k+1})\end{aligned}$$

We find that the adjoint of the symplectic Euler scheme is also explicit. Obviously, the adjoint of the adjoint of a flow map is simply the flow map itself.

3.4 Building Symplectic Integrators: Splitting Methods

Let $H(\mathbf{q}, \mathbf{p}) = H_1(\mathbf{q}, \mathbf{p}) + H_2(\mathbf{q}, \mathbf{p})$ have flow map Φ_h, and let Φ_h^1, Φ_h^2 be the flow maps for the systems with Hamiltonians H_1, H_2 respectively. We propose that the map

$$\Psi_h \coloneqq \Phi_h^1 \circ \Phi_h^2$$

is an approximation of Φ_h. For this to be a first order method, we need at least

$$||\Psi_h(\mathbf{u}) - \Phi_h(\mathbf{u})|| \leq C(\mathbf{u})h^2$$

That is, the local error must be 2nd order. We expand the exact flow $\Phi_h(\mathbf{u})$ as a Taylor series (remember that we are expanding it as a function of t, so all derivatives are with respect to t):

$$\begin{aligned}\Phi_h(\mathbf{u}) &= \mathbf{u} + h\big[\Phi_h(\mathbf{u})\big]' + \mathcal{O}(h^2)\\ &= \mathbf{u} + h\mathbf{J}\nabla H(\mathbf{u}) + \mathcal{O}(h^2)\\ &= \mathbf{u} + h(\mathbf{J}\nabla H_1 + \mathbf{J}\nabla H_2)(\mathbf{u}) + \mathcal{O}(h^2)\end{aligned} \tag{3.1}$$

On the other hand, we have

$$\begin{aligned}\Phi_h^1(\mathbf{u}) &= \mathbf{u} + h\mathbf{J}\nabla H_1(\mathbf{u}) + \mathcal{O}(h^2)\\ \Phi_h^2(\mathbf{u}) &= \mathbf{u} + h\mathbf{J}\nabla H_2(\mathbf{u}) + \mathcal{O}(h^2)\end{aligned}$$

and composing them gives the following (remember that $\nabla H_1, \nabla H_2$ are simply Jacobian matrices, with $\nabla H_1(\mathbf{u}), \nabla H_2(\mathbf{u})$ vectors evaluated at $\mathbf{u}$.

$$\begin{aligned}\Phi_h^1 \circ \Phi_h^2 &= \mathbf{u} + h\mathbf{J}\nabla H_2(\mathbf{u}) + h\mathbf{J}\nabla H_1\big(\mathbf{u} + h\mathbf{J}\nabla H_2(\mathbf{u})\big) + \mathcal{O}(h^2)\\ &= \mathbf{u} + h\mathbf{J}\nabla H_2(\mathbf{u}) + h\mathbf{J}\nabla H_1(\mathbf{u}) + \underbrace{h^2(\mathbf{J}\nabla H_1) \circ (\mathbf{J}\nabla H_2)(\mathbf{u})\big)}_{\mathcal{O}(h^2)} + \mathcal{O}(h^2)\\ &= \mathbf{u} + h(\mathbf{J}\nabla H_2 + \mathbf{J}\nabla H_1)(\mathbf{u}) + \mathcal{O}(h^2)\end{aligned}$$

which agrees with the terms of (3.1) up to second order, and therefore the local error is indeed second order.

3.4.1 Symplectic Euler Constructed from Splitting Schemes

Let $H_1(\mathbf{q}, \mathbf{p}) = \mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}/2$ and $H_2(\mathbf{q}, \mathbf{p}) = U(\mathbf{q})$, then the splitting method for $H = H_1 + H_2$ can be obtained by determining the flow maps for each of the two parts. For H_1 and H_2 we have the differential equations

$$\begin{cases}\dot{\boldsymbol{q}} = \mathbf{M}^{-1}\mathbf{p}\\ \dot{\boldsymbol{p}} = -\nabla_{\mathbf{q}}U(\mathbf{q})\end{cases} \implies H_1 \begin{cases}\dot{\boldsymbol{q}} = \mathbf{M}^{-1}\mathbf{p}\\ \dot{\boldsymbol{p}} = \mathbf{0}\end{cases} \quad \text{and } H_2 \begin{cases}\dot{\boldsymbol{q}} = \mathbf{0}\\ \dot{\boldsymbol{p}} = -\nabla_{\mathbf{q}}U(\mathbf{q})\end{cases}$$

The fact that $\dot{\boldsymbol{p}} = 0$ for H_1 tells us that the momentum is constant and therefore the trajectory $\mathbf{q}$ is linear, and hence the flow map (integration) Φ_h^1 is

$$\Phi_h^1 \begin{pmatrix} \mathbf{q}_k \\ \mathbf{p}_k \end{pmatrix} = \begin{pmatrix} \mathbf{q}_{k+1} \\ \mathbf{p}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_k \\ \mathbf{p}_k \end{pmatrix}$$

The flow map Φ_h^2 is

$$\Phi_h^2 \begin{pmatrix} \mathbf{q}_k \\ \mathbf{p}_k \end{pmatrix} = \begin{pmatrix} \mathbf{q}_{k+1} \\ \mathbf{p}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{q}_k \\ \mathbf{p}_k - h\nabla U(\mathbf{q}_k) \end{pmatrix}$$

The composition of these maps is

$$\Phi_h^1 \circ \Phi_h^2 = \begin{cases} \mathbf{q}_{k+1} = \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_{k+1} \\ \mathbf{p}_{k+1} = \mathbf{p}_k - h\nabla U(\mathbf{q}_k) \end{cases}$$

which is precisely the symplectic Euler method. Composing the same two maps in the opposite order gives the adjoint symplectic Euler method.

$$\left(\Phi_h^1 \circ \Phi_h^2\right)^\dagger = \Phi_h^2 \circ \Phi_h^1 = \begin{cases} \mathbf{q}_{k+1} = \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_k \\ \mathbf{p}_{k+1} = \mathbf{p}_k - h\nabla U(\mathbf{q}_{k+1}) \end{cases}$$

3.4.2 Symplectic Verlet Method from Splitting Schemes

For the symplectic Euler method Φ_h and its adjoint method $\Phi_h^\dagger$, consider the composition

$$\mathcal{K}_h \coloneqq \Phi_{h/2}^\dagger \circ \Phi_{h/2}$$

Computing this, we have $\Phi_{h/2}(\mathbf{q}_k, \mathbf{p}_k) = (\mathbf{q}_{k+1/2}, \mathbf{p}_{k+1/2})$, and $\Phi_{h/2}^\dagger(\mathbf{q}_{k+1/2}, \mathbf{p}_{k+1/2}) = (\mathbf{q}_{k+1}, \mathbf{p}_{k+1})$ defined

$$\Phi_{h/2} \begin{pmatrix} \mathbf{q}_k \\ \mathbf{p}_k \end{pmatrix} = \begin{cases} \mathbf{q}_{k+1/2} = \mathbf{q}_k + \frac{h}{2}\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\ \mathbf{p}_{k+1/2} = \mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k) \end{cases}$$

$$\Phi_{h/2}^\dagger \begin{pmatrix} \mathbf{q}_{k+1/2} \\ \mathbf{p}_{k+1/2} \end{pmatrix} = \begin{cases} \mathbf{q}_{k+1} = \mathbf{q}_{k+1/2} + \frac{h}{2}\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\ \mathbf{p}_{k+1} = \mathbf{p}_{k+1/2} - \frac{h}{2}\nabla U(\mathbf{q}_{k+1}) \end{cases}$$

This composition simplifies to

$$\mathcal{K}_h \coloneqq \Phi_{h/2}^\dagger \circ \Phi_{h/2} = \begin{cases} \mathbf{p}_{k+1/2} & = \mathbf{p}_k - \frac{h}{2}\nabla U(\mathbf{q}_k) \\ \mathbf{q}_{k+1} & = \mathbf{q}_k + h\mathbf{M}^{-1}\mathbf{p}_{k+1/2} \\ \mathbf{p}_{k+1} & = \mathbf{q}_{k+1/2} - \frac{h}{2}\nabla U(\mathbf{q}_{k+1}) \end{cases}$$

which is precisely the velocity Verlet method in Hamiltonian form. Since we have obtained this method as the composition of two symplectic maps, and the symplectic maps form a group, we know that this method will also be symplectic. Similarly, we can construct the adjoint map of $\mathcal{K}_h$ by simply taking the composition in the other direction, which we see to be the same (i.e. $\mathcal{K}_h$ is symmetric/self-adjoint).

$$\mathcal{K}_h^\dagger \coloneqq \left(\Phi_{h/2}^\dagger \circ \Phi_{h/2}\right)^\dagger = \Phi_{h/2}^\dagger \circ \Phi_{h/2} = \mathcal{K}_h$$

3.4.3 General Composition Methods

In general, if we have any two symplectic numerical methods, say $\hat{\Phi}_h^1$ and $\hat{\Phi}_h^2$, then the composition

$$\hat{\Phi}_h \coloneqq \hat{\Phi}_h^1 \circ \hat{\Phi}_h^2$$

is another symplectic numerical method. The order of this new method is typically the minimum of the orders of the two methods involved, but it can be higher, as the example of the Verlet method (constructed by composing the Euler and its adjoint, both of order 1).

3.4.4 Yoshida Fourth Order Integration

The Yoshida Fourth Order Scheme is overall three iterations of velocity Verlet (making it also a symplectic integrator), using stepsizes $\tau_0 h, \tau_1 h, \tau_0 h$, respectively. We write this with subindices α, β to indicate the intermediate stages, and abuse our notation to be similar to those in computer science.

$$\begin{aligned}
\mathbf{p}_\alpha &= \mathbf{p} - (\tau_0\, h/2)\nabla U(\mathbf{q}) \\
\mathbf{q}_\alpha &= \mathbf{q} + (\tau_0\, h)\mathbf{M}^{-1}\mathbf{p}_\alpha \\
\mathbf{p}_\alpha &= \mathbf{p}_\alpha - (\tau_0\, h/2)\nabla U(\mathbf{q}_\alpha) \\
\mathbf{p}_\beta &= \mathbf{p}_\alpha - (\tau_1\, h/2)\nabla U(\mathbf{q}_\alpha) \\
\mathbf{q}_\beta &= \mathbf{q}_\alpha + (\tau_1\, h)\mathbf{M}^{-1}\mathbf{p}_\beta \\
\mathbf{p}_\beta &= \mathbf{p}_\beta - (\tau_1\, h/2)\nabla U(\mathbf{q}_\beta) \\
\mathbf{p} &= \mathbf{p}_\beta - (\tau_0\, h/2)\nabla U(\mathbf{q}_\beta) \\
\mathbf{q} &= \mathbf{q}_\beta + (\tau_0\, h)\mathbf{M}^{-1}\mathbf{p} \\
\mathbf{p} &= \mathbf{p} - (\tau_0\, h/2)\nabla U(\mathbf{q})
\end{aligned}$$

The equations can be written in a simplified form, combining several of the steps. This scheme requires three new evaluations of the force ∇U per iteration, making it significantly more expensive than the vanilla second-order Verlet method.

4 Analyzing Geometric Integrators

5 Potential Energy Surfaces

In particle physics and other contexts, we wish to consider systems of indistinguishable particles $\mathbf{r}_j \in \mathbb{R}^3$. A **configuration** of J particles is an mset $\{\mathbf{r}_1, \ldots, \mathbf{r}_J\} \subset \mathbb{R}^3$, the set of all configurations of J particles is denoted

$$\mathcal{R}_J \coloneqq \big\{\{\mathbf{r}_1, \ldots, \mathbf{r}_J\}\big\}$$

and the set of all admissible finite configurations is given by

$$\mathcal{R} \coloneqq \bigcup_{J=0}^{\infty} \mathcal{R}_J$$

A **potential energy surface (PES)** is a mapping $E : \mathcal{R} \longrightarrow \mathbb{R}$ that represents the potential energy of a system of an arbitrary number of particles in a certain configuration. Clearly, it should be invariant under translations, permutations, and isometries. That is, given a configuration $R \coloneqq \{\mathbf{r}_j\}_j \subset \mathbb{R}^3$, the invariance can be written

$$E\big(\{Q\mathbf{r}_{\sigma j}\}_j\big) = E\big(\{\mathbf{r}_j\}_j\big) \text{ for all } Q \in O(3), \sigma \text{ a permutation}$$

The simplest and most widely studied potential function is the **Lennard-Jones potential**. Given two particles $\mathbf{r}_1$ and $\mathbf{r}_2$, let $r = ||\mathbf{r}_1 - \mathbf{r}_2||$ be the distance between them. Then, the LJ potential is evaluated as a function of the distance between the two particles, with σ some known constant dependent on what particle we are modelling.

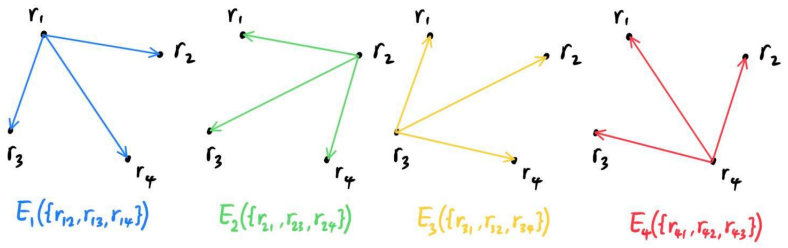
$$V_{\mathrm{LJ}}(\mathbf{r}_1, \mathbf{r}_2) = V_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Clearly, V_{LJ} satisfies translation, permutation, and isometry invariance. All interatomic potential models make various assumptions on the PES regarding low-rank structures. In general, we can

represent a complex fully many-body PES E as a combination of simple components E_i that takes in position vectors relative to some ith particle. Hence, we assume the following decomposition.

$$E(\{\mathbf{r}_1, \ldots, \mathbf{r}_J\}) = \sum_{i=1}^{J} E_i(\{\mathbf{r}_j - \mathbf{r}_i\}_{j \neq i})$$

Here, we sum over i the J center atoms, which automatically makes E translation invariant. For $J = 4$, we can visualize the decomposition as



Furthermore, each component function E_i can be decomposed into a sum of functions V_N, called **N-body functions**, which depend on the N differences to the center atom i.

$$E_i\big(\{\mathbf{r}_{ij}\}_{j=1}^{J-1}\big) = V_0 + \sum_{N=1}^{\mathcal{N}} \sum_{j_1 < j_2 < \ldots < j_N} V_N(\mathbf{r}_{ij_1}, \ldots, \mathbf{r}_{ij_N})$$

$\mathcal{N}$ is called the **maximal order of interaction**, which represents the truncation of the expansion, and in full generality, we would have $\mathcal{N} = J - 1$. Other values of $\mathcal{N}$ would reduce the V_N's to 0 for $N > \mathcal{N}$. Taking a look at this decomposition for E_1 of the 4-body problem

$$E_1(\{\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{14}\}) = V_0 + \sum_{j_1} V_1(\mathbf{r}_{1j_1}) + \sum_{j_1 < j_2} V_2(\mathbf{r}_{1j_1}, \mathbf{r}_{1j_2}) + \sum_{j_1 < j_2 < j_3} V_3(\mathbf{r}_{1j_1}, \mathbf{r}_{1j_2}, \mathbf{r}_{1j_3})$$

we have the following:

$$\begin{aligned}
V_0 &= V_0 \\
\sum_{j_1} V_1(\mathbf{r}_{1j_1}) &= V_1(\mathbf{r}_{12}) + V_1(\mathbf{r}_{13}) + V_1(\mathbf{r}_{14}) \\
\sum_{j_1 < j_2} V_2(\mathbf{r}_{1j_1}, \mathbf{r}_{1j_2}) &= V_2(\mathbf{r}_{12}, \mathbf{r}_{13}) + V_2(\mathbf{r}_{12}, \mathbf{r}_{14}) + V_2(\mathbf{r}_{13}, \mathbf{r}_{14}) \\
\sum_{j_1 < j_2 < j_3} V_3(\mathbf{r}_{1j_1}, \mathbf{r}_{1j_2}, \mathbf{r}_{1j_3}) &= V_3(\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{14})
\end{aligned}$$

If we have $\mathcal{N} = 2$, then the maximal order of interaction would be 2, and V_3 would vanish, leaving us with

$$\begin{aligned}
E_1(\{\mathbf{r}_{12}, \mathbf{r}_{13}, \mathbf{r}_{14}\}) = V_0 &+ V_1(\mathbf{r}_{12}) + V_1(\mathbf{r}_{13}) + V_1(\mathbf{r}_{14}) \\
&+ V_2(\mathbf{r}_{12}, \mathbf{r}_{13}) + V_2(\mathbf{r}_{12}, \mathbf{r}_{14}) + V_2(\mathbf{r}_{13}, \mathbf{r}_{14})
\end{aligned}$$

Note that the arguments in V_N are taken as (vectoral) differences with respect to a center atom, but for simplicity, we still denote them by $R = (\mathbf{r}_1, \ldots, \mathbf{r}_N) \in \mathbb{R}^{3N}$. Moreover, we treat the V_N as functions on $\mathbb{R}^{3N}$ rather than on configurations. We further assume that the V_N satisfy the following conditions

1. Regularity. $V_N \in C^t(\mathbb{R}^{3N} \setminus \{0\})$, where $t \geq 1$.

2. Locality. There exists $r_{\text{cut}} > 0$ such that $V_N(R) = 0$ if

$$\max_{1 \le j \le N} |\mathbf{r}_j| \ge r_{\text{cut}}$$

That is, if at least one of the distances of the N particles (from a center particle $\mathbf{r}_i$) surpass r_{cut}, then V_N vanishes.

3. Isometry and Permutation Invariance. For all $Q \in \mathrm{O}(3)$, $\sigma \in S_N$, we have $V_N(QR) = V_N(R)$ and $V_N(\sigma R) = V_N(R)$.

Since N-body potentials typically have a singularity at $r = 0$, we will not study convergence up to collision but instead specify some minimal distance $r_0 > 0$. Therefore, we can define the domains

$$\Omega_{r_0} \coloneqq \{\mathbf{r} \in \mathbb{R}^3 \,|\, |\mathbf{r}| > r_0\}$$
$$\Omega_{r_0}^N = \{(\mathbf{r}_1, \dots, \mathbf{r}_N) \in \mathbb{R}^{3N} \mid \min_j r_j > 0\}$$

We can imagine $\Omega_{r_0}^N$ as $\mathbb{R}^{3N}$ with a closed $3N$-ball of radius r_0 at the origin subtracted out. The associated function spaces would be

$$C^s_{r_{\text{cut}}}(r_0, \infty) \coloneqq \left\{f \in C^s(r_0, \infty) \,|\, f = 0 \text{ in } [r_{\text{cut}}, \infty)\right\}$$
$$C^s_{r_{\text{cut}}}(\Omega^N_{r_0}) \coloneqq \left\{f \in C^s(\Omega^N_{r_0}) \,|\, f(\mathbf{r}_1, \dots, \mathbf{r}_N) = 0 \text{ if } r_j \ge r_{\text{cut}} \text{ for any } j\right\}$$

Note that the functions we are interested in aren't even defined for when any of the distances are less than r_0 and are 0 when any are more than r_{cut}. Note that the standard 2-norm of a function f of one real variable is

$$||f||_2 \coloneqq \left(\int_{\mathbb{R}} |f(x)|^2 \, dx\right)^{1/2}$$

and in general, that of a function $F : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ is

$$||F||_2 \coloneqq \left(\int_{\mathbb{R}^n} ||F(\mathbf{x})||^2 \, d\mathbf{x}\right)^{1/2}$$

where $||\cdot||$ is a norm in $\mathbb{R}^p$.

6 Spherical Harmonics

Spherical harmonics are a set of functions used to represent functions on the surface of a sphere $\mathbb{S}^2$. An analogue is the **fourier series**, which is a sum that represents a periodic function as a sum of sine and cosine waves. Note that the period of $\sin nx$ and $\cos nx$, where n is a positive integer, is $2\pi/n$. Therefore, let $f(x)$ be defined in the interval $(-L, L)$ and outside of this interval by $f(x+2L) = f(x)$ (i.e. $f(x)$ is $2L$-periodic. Then, the **Fourier expansion** corresponding to $f(x)$ is given by

$$f(x) \approx \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L}\right)$$

where the Fourier coefficients a_n, b_n are for $n = 0, 1, 2, \dots$

$$a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} \, dx$$
$$b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} \, dx$$

Similarly, given a point in terms of the spherical coordinate system,

$$\mathbf{r} = r\hat{\boldsymbol{r}} = r(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$$

the pair (ϕ, θ) paramaterizes the unit vector $\hat{\boldsymbol{r}} \in \mathbb{S}^2 \subset \mathbb{R}^3$. We can define the complex spherical harmonics $Y_l^m : \mathbb{S}^2 \longrightarrow \mathbb{C}$ indexed by $l \in \mathbb{N}, m = -l, \dots l$ given by

$$Y_l^m(\hat{\boldsymbol{r}}) \coloneqq P_l^m(\cos\theta)\, e^{mi\phi}$$

where P_l^m is a normalization factor (associated with the Legendre polynomials P_l^m).

6.0.1 Legendre Polynomials

The **(ordinary) Legendre polynomials** are a system of complete and orthogonal polynomials, defined in many ways. Here is the simplest way to define them: Over the interval $[-1, 1]$, we can define $P_n(x)$ as a polynomial of degree n such that

$$\int_{-1}^{1} P_m(x)\, P_n(x)\, dx = 0 \text{ if } n \neq m$$

which determines the polynomials completely up to an overall scale factor, which is fixed by the standardization that $P_n(1) = 1$. We can define the P_n's recursively.

1. $P_0(x) = 1$ from standardization property.
2. $P_1(x)$ must be orthogonal to $P_0(x)$, leaving only the chose of $P_1(x) = x$.
3. P_2 must be orthogonal to P_0 and P_1, and we can continue to find the coefficients as such: $P_2(x) = \frac{1}{2}(3x^2 - 1)$
4. $P_3(x) = \frac{1}{2}(5x^3 - 3x)$.
5. $P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$
6. etc.

Given the function space $C([-1, 1])$ of functions mapping from $[-1, 1]$ to $\mathbb{R}$, we can define an inner product to be

$$\langle f, g \rangle \coloneqq \int_{-1}^{1} f(x)\, g(x)\, dx$$

i.e. the functions f and g are orthogonal when this integral is 0. Indeed we can check that that Legendre polynomials are orthogonal, since for instance,

$$\int_{-1}^{1} P_2(x)\, P_3(x)\, dx = \int_{-1}^{1} \frac{1}{4}(3x^2 - 1)(5x^3 - 3x)\, dx = 0$$

Furthermore, this inner product induces the L2 norm on the function space, defined

$$||f||_2 \coloneqq \langle f, f \rangle^{\frac{1}{2}} = \left(\int_{-1}^{1} f(x)^2\, dx \right)^{\frac{1}{2}}$$

If one begins with the monomial sequence $\{1, x, x^2, \dots\}$ on the interval $[-1, 1]$ and applies the Gram-Schmidt process, then one obtains the Legendre polynomials.

Now, a related but not equivalent set is the **associated Legendre polynomials**, which are the canonical solutions to the general Legendre equation

$$\left(1 - x^2\right) \frac{d^2}{dx^2} P_l^m(x) - 2x \frac{d}{dx} P_l^m(x) + \left[l(l+1) - \frac{m^2}{1 - x^2} \right] P_l^m(x) = 0$$

Its simplest definition is in terms of derivatives of ordinary Legendre polynomials

$$\begin{aligned} P_l^m(x) &= (-1)^m (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_l(x) \\ &= \frac{(-1)^m}{2^l l!} (1 - x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} (x^2 - 1)^l \end{aligned}$$

which is defined for all $l \in \mathbb{N}, -l \leq z \leq l$. However, they are nor mutually orthogonal in general. For example, P_1^1 is not orthogonal to P_2^2. However, if either the m index or the l index is fixed, then they are orthogonal.

$$\langle P_k^m, P_l^m \rangle = 0 \text{ and } \langle P_l^m, P_l^n \rangle = 0$$

As we see in our definition of spherical harmonics, we can reparameterize them in terms of angles, letting $x = \cos\theta$. Then,

$$P_l^m(\cos\theta) = (-1)^m (\sin\theta)^m \frac{d^m}{d(\cos\theta)^m} P_l(\cos\theta)$$

6.0.2 Transformations of Spherical Harmonics

Definition (L^2 Function Space) Let X be a measurable space. The set $L^2(X)$ consists of all functions $f : X \longrightarrow \mathbb{C}$ where

$$\int_X |f(x)|^2 \, dx < +\infty$$

The spherical harmonics $Y_l^m : \mathbb{S}^2 \longrightarrow \mathbb{C}$ form a complete set of orthonormal functions and thus forms an orthonormal basis of the Hilbert space $L^2(\mathbb{S}^2)$ (this is an important fact). One the unit sphere $\mathbb{S}^2$, and square-integrable function $f : \mathbb{S}^2 \longrightarrow \mathbb{C}$ can be expanded as a linear combination of these a

$$f(\hat{\mathbf{r}}) = f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_l^m Y_l^m(\theta, \phi)$$

Rotations $Y_l^m(Q\mathbf{r})$ can be conveniently expressed and manipulated using Wigner D-matrices, which we will introduce now.

Definition (Wigner D-Matrices) Let Y_l^m be a given complex spherical harmonic and $Q \in \mathrm{SO}(3)$. Then, there exists $D_{\mu m}^l(Q) \in \mathbb{C}$ with $\mu, m = -l, \ldots, l$ (each of the $(2l+1)(2l+1)$ terms measurable with respect to the normalized Haar measure on $\mathrm{SO}(3)$), such that

$$Y_l^m(Q\hat{\mathbf{r}}) = \sum_{\mu=-l}^{l} D_{\mu m}^l(Q) \, Y_l^\mu(\hat{\mathbf{r}}) \tag{D}$$

We can express this in matrix form. Given the set of $2l+1$ complex spherical harmonics Y_m^l's with $l \in \mathbb{N}$ fixed and $m = -l, \ldots, l$, a matrix $D^l(Q)$ exists such that

$$\begin{pmatrix} Y_l^{m=-l}(Q\hat{\mathbf{r}}) \\ \vdots \\ Y_l^{m=l}(Q\hat{\mathbf{r}}) \end{pmatrix} = \begin{pmatrix} D_{\mu=-l,m=-l}^l(Q) & \cdots & D_{\mu=l,m=-l}^l(Q) \\ \vdots & \ddots & \vdots \\ D_{\mu=-l,m=l}^l(Q) & \cdots & D_{\mu=l,m=l}^l(Q) \end{pmatrix} \begin{pmatrix} Y_l^{\mu=-l}(\hat{\mathbf{r}}) \\ \vdots \\ Y_l^{\mu=l}(\hat{\mathbf{r}}) \end{pmatrix}$$

Remember, that l is fixed, and μ, m represents the indices.

In summary, the spherical harmonics exhibit a nice property that a rotation of the inputs can be simply modeled as a linear combination of spherical harmonics. Note that for scalar indices $\mu, m \in \{-(2l+1), \ldots, 2l+1\}$, $D^l(Q)$ is a $(2l+1) \times (2l+1)$ matrix representing the rotation of a single particle $\hat{\mathbf{r}}$. But when we are talking about the system of particles $\hat{\mathbf{R}}$, we are dealing not just with a single harmonic Y_l^m but a tensor product of harmonics

$$Y_{\mathbf{l}}^{\mathbf{m}} = Y_{l_1}^{m_1} \otimes Y_{l_2}^{m_2} \otimes \ldots \otimes Y_{l_N}^{m_N}$$

Then the Wigner D-matrix $D^{\mathbf{l}}(Q) = D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}$ is indexed with vectors $\boldsymbol{\mu}, \mathbf{m}$, with shape

$$\left[\prod_{\alpha=1}^{N}(2l_\alpha+1)\right] \times \left[\prod_{\alpha=1}^{N}(2l_\alpha+1)\right]$$

resulting in

$$\begin{aligned}
Y^{\mathbf{m}}_{\mathbf{l}}(Q\hat{\mathbf{R}}) &\coloneqq Y^{m_1}_{l_1}(Q\hat{\mathbf{r}}_1) \otimes Y^{m_2}_{l_2}(Q\hat{\mathbf{r}}_2) \otimes \ldots \otimes Y^{m_N}_{l_N}(Q\hat{\mathbf{r}}_N) \\
&= \left(\sum_{\mu_1=-l_1}^{l_1} D^{l_1}_{\mu_1 m_1}(Q) Y^{\mu_1}_{l_1}(\hat{\mathbf{r}}_1)\right) \otimes \ldots \otimes \left(\sum_{\mu_N=-l_N}^{l_N} D^{l_N}_{\mu_N m_N}(Q) Y^{\mu_N}_{l_N}(\hat{\mathbf{r}}_N)\right) \\
&= \sum_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}} \bigotimes_{i=1}^{N} D^{l_i}_{\mu_i m_i}(Q)\, Y^{\mu_i}_{l_i}(\hat{\mathbf{r}}_i) \\
&= \sum_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}} \left[\prod_{i=1}^{N} D^{l_i}_{\mu_i m_i}(Q)\right] \left(Y^{\mu_1}_{l_1} \otimes \ldots \otimes Y^{\mu_N}_{l_N}\right) \\
&= \sum_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, Y^{\boldsymbol{\mu}}_{\mathbf{l}}(\hat{\mathbf{R}}) \qquad \text{with } D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) = \prod_{\alpha=1}^{N} D^{l_\alpha}_{\mu_\alpha m_\alpha}(Q)
\end{aligned} \tag{1}$$

where the fact that

$$D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) = \prod_{\alpha=1}^{N} D^{l_\alpha}_{\mu_\alpha m_\alpha}(Q)$$

follows naturally since we distribute the tensor products over addition and gather the coefficients. As an example, let us have $N = 3$ with $\mathbf{l} = (2, 1, 3)$, with $\mathbf{m} = (-2, 1, 1)$. Then,

$$\begin{aligned}
Y^{\mathbf{m}}_{\mathbf{l}} &\coloneqq Y^{-2}_{2}(Q\hat{\mathbf{r}}_1) \otimes Y^{1}_{1}(Q\hat{\mathbf{r}}_2) \otimes Y^{1}_{3}(Q\hat{\mathbf{r}}_3) \\
&= \left(\sum_{\mu_1=-2}^{2} D^{2}_{\mu_1,-2}(Q) Y^{\mu_1}_{2}(\hat{\mathbf{r}}_1)\right) \otimes \left(\sum_{\mu_2=-1}^{1} D^{1}_{\mu_2,1}(Q) Y^{\mu_2}_{1}(\hat{\mathbf{r}}_2)\right) \otimes \left(\sum_{\mu_3=-3}^{3} D^{3}_{\mu_3,1}(Q) Y^{\mu_3}_{3}(\hat{\mathbf{r}}_3)\right)
\end{aligned}$$

which requires evaluations of the Wigner D-matrices $D^2(Q) \in \mathbb{R}^{5\times5}, D^1(Q) \in \mathbb{R}^{3\times3}, D^3(Q) \in \mathbb{R}^{7\times7}$. The reason we take the trouble to choose spherical harmonics at all is because rotations can be easily represented as linear combinations. That is,

$$Y^{\mathbf{m}}_{\mathbf{l}} \circ Q = \sum_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, Y^{\boldsymbol{\mu}}_{\mathbf{l}}$$

but since

$$\begin{aligned}
\phi_{\mathbf{nlm}}(\mathbf{R}) &= \prod_{\alpha=1}^{N} \phi_{n_\alpha l_\alpha m_\alpha}(\mathbf{r}_j) \\
&= \prod_{\alpha=1}^{N} P_{n_\alpha}(r_\alpha) Y^{m_\alpha}_{l_\alpha}(\hat{\mathbf{r}}_\alpha)
\end{aligned}$$

and the radial components $P_{n_\alpha}(r_\alpha)$ are already rotation-invariant, this implies that

$$\phi_{\mathbf{nlm}} \circ Q = \sum_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, \phi_{\mathbf{nlm}}$$

This is extremely useful when integrating over O(3) since computing the left hand integral is difficult, since the composition is too abstract to deal with. However, linear combinations are easily to integrate over.

Definition (Clebsch-Gordan Coefficients) The **Clebsch-Gordan coefficients** are the coefficients appearing in the expansion of the product of two spherical harmonics in terms of spherical harmonics themselves. That is, given $\lambda, -\lambda \leq \mu \leq \lambda$ with spherical harmonic Y_λ^μ, there exists coefficients $C^{\lambda\mu}_{l_1 m_1 l_2 m_2} \in \mathbb{R}$, with $-l_i \leq m_i \leq l_i$, $|l_1 - l_2| \leq \lambda \leq l_1 + l_2$, $-\lambda \leq \mu \leq \lambda$ such that

$$Y_\lambda^\mu = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C^{\lambda\mu}_{l_1 m_1 l_2 m_2} Y_{l_1}^{m_1} Y_{l_2}^{m_2}$$

where $C^{\lambda\mu}_{l_1 m_1 l_2 m_2} = 0$ if $m_1 + m_2 \neq \mu$. That is, we can expand Y_λ^μ as a linear combination of lower-degree harmonics satisfying the above properties.

7 Approximation by Tensor Products

Given a multivariate function $f : \mathbb{R}^{3N} \longrightarrow \mathbb{R}$, let us attempt to approximate it using a tensor product basis. Let us have

1. $\mathbf{n} = (n_1, n_2, \ldots, n_N) \in \mathbb{N}_0^N$
2. $\mathbf{l} = (l_1, l_2, \ldots, l_N) \in \mathbb{N}_0^N$
3. $\mathbf{m} = (m_1, m_2, \ldots, m_N) \in \mathcal{M}_\mathbf{l} \coloneqq \{\boldsymbol{\mu} \in \mathbb{Z}^N \mid -l_\alpha \leq \mu_\alpha \leq l_\alpha\}$. This is simple to understand. If we have $N = 3$ with $\mathbf{l} = (1, 0, 1)$, then the three possible choices for $\mathbf{m}$ are

$$\mathbf{m} = \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\-1\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}, \begin{pmatrix}-1\\0\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\-1\end{pmatrix}$$

Remember that given two functions $f, g : X \longrightarrow \mathbb{R}$, $f \otimes g : X \times X \longrightarrow \mathbb{R}$, with

$$(f \otimes g)(x, y) \coloneqq f(x) \cdot g(y)$$

where $\cdot$ is regular multiplication in $\mathbb{R}$. Therefore, given the vectors $\mathbf{n}, \mathbf{l}, \mathbf{m}$ above, we can approximate

$$f : \mathbb{R}^{3N} = \underbrace{\mathbb{R}^3 \times \ldots \times \mathbb{R}^3}_{N \text{ times}} \longrightarrow \mathbb{R}$$

as a tensor product of functions $\phi_{\mathbf{nlm}} = \phi_{n_1 l_1 m_1} \otimes \phi_{n_2 l_2 m_2} \otimes \ldots \otimes \phi_{n_N l_N m_N}$, where each $\phi_{n_j l_j m_j} : \mathbb{R}^3 \longrightarrow \mathbb{R}$ is defined

$$\phi_{nlm} \coloneqq P_n(r)\, Y_l^m(\hat{\mathbf{r}})$$

Therefore, we have the following approximation

$$\begin{aligned}\phi_{\mathbf{nlm}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) &= \left(\bigotimes_{j=1}^{N} \phi_{n_j l_j m_j}\right)(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \\ &= \prod_{j=1}^{N} \phi_{n_j l_j m_j}(\mathbf{r}_j)\end{aligned}$$

For example, let $N = 4$ and

$$\mathbf{n} = \begin{pmatrix}3\\4\\2\\8\end{pmatrix} \in \mathbb{N}_0^4, \mathbf{l} = \begin{pmatrix}2\\0\\0\\1\end{pmatrix} \in \mathbb{N}_0^4, \mathbf{m} = \begin{pmatrix}-2\\0\\0\\-1\end{pmatrix} \in \mathcal{M}_\mathbf{l} \subset \mathbb{Z}^4$$

Then

$$\begin{aligned}\phi_{\mathbf{nlm}}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) &= \prod_{j=1}^{4} \phi_{n_j l_j m_j}(\mathbf{r}_j) \\ &= \phi_{3,2,-2}(\mathbf{r}_1) \cdot \phi_{4,0,0}(\mathbf{r}_2) \cdot \phi_{2,0,0}(\mathbf{r}_3) \cdot \phi_{8,1,-1}(\mathbf{r}_4) \\ &= P_3(r_1) Y_2^{-2}(\hat{\boldsymbol{r}}_1) \cdot P_4(r_2) Y_0^0(\hat{\boldsymbol{r}}_2) \cdot P_2(r_3) Y_0^0(\hat{\boldsymbol{r}}_3) \cdot P_8(r_4) Y_1^{-1}(\hat{\boldsymbol{r}}_4)\end{aligned}$$

Therefore, the (countably infinite) set

$$\Phi_N \coloneqq \{\phi_{\mathbf{nlm}} \,|\, \mathbf{n}, \mathbf{l} \in \mathbb{N}^N, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}\}$$

is a tensor product basis, but this is known to be too large to work with. However, it has a very nice property that is proved in the original ACE paper:

Proposition (Approximation Using Tensor Product Basis) Any function $f \in C^s_{r_{\text{cut}}}(\Omega^N_{r_0})$ can be approximated arbitrarily well by linear combinations of the basis functions, which is proven by showing that the product basis $\Phi_N = \{\phi_{\mathbf{nlm}}\}$ is dense in

$$\bigotimes_{i=1}^{N} Y$$

where $Y = \{f \in C^s(\overline{\Omega}_{r_0}) \,|\, f = 0 \text{ in } B^c_{r_{\text{cut}}}\}$ is some function f defined in $\overline{\Omega}_{r_0} = \{\mathbf{r} \in \mathbb{R}^3 \,|\, ||\mathbf{r}|| > r_0\}$ that vanishes for values outside the ball with radius r_{cut}.

8 Symmetric Polynomials

We have scene that we can approximate an N-body function V_N from the tensor product space $\text{span}\{\phi_{\mathbf{nlm}}\}$.

1. We will now show that if V_N satisfies the permutation and isometry symmetries, then the approximant may be assumed to inherit these symmetries without loss of accuracy.
2. We will construct an explicit basis that respects these symmetries.

There is a very general strategy to construct a basis respecting the desired symmetries at least in principle.

1. Define the normalized Haar measure H on the compact group $G \coloneqq \langle S_N \cup \text{O}(3) \rangle$ (generator) obtained by joining the permutation and isometry groups.
2. Compute the symmetrised functions

   $$\phi^{\text{sym}}_{\mathbf{nlm}}(R) \coloneqq \int_{g \in G} \phi_{\mathbf{nlm}}(gR) \; H(dg)$$

   which may now be linearly dependent.
3. Construct a basis of

   $$\text{span}\big\{\phi^{\text{sym}}_{\mathbf{nlm}} \,|\, \mathbf{n}, \mathbf{l} \in \mathbb{N}^N, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}\big\}$$

Notice that step 2 is easy to do analytically, while step 3 is not so obvious and is usually done numerically. It is also computationally expensive.

Proposition (Accuracy of Symmetric Basis) Let G be a group acting on $\mathbb{R}^d$, $\Omega \subset \mathbb{R}^d$ invariant under G, and X a subspace of functions $f : \Omega \longrightarrow \mathbb{R}$ such that $f \circ g = f$ for all $g \in G$. Further, let a norm $||\cdot|| : X \longrightarrow \mathbb{R}$ be invariant under G, i.e.

$$||f|| = ||f \circ g|| \text{ for all } g \in G$$

Next, let $\tilde{f} \in X$ be some approximation to f, not necessarily respecting the G symmetries, and let

$$\tilde{f}^{\rm sym} \coloneqq \int_G \tilde{f} \circ g \; H(dg)$$

be the relevant symmetrized function. Then, since f is invariant under $g \in G$, we have

$$\begin{aligned} ||f - \tilde{f}^{\rm sym}|| &= \left|\left| \int_G (f - \tilde{f}) \circ g \; H(dg) \right|\right| \\ &\le \int_G ||(f - \tilde{f} \circ g|| \; H(dg) \\ &= \int_G ||f - \tilde{f}|| \; H(dg) = ||f - \tilde{f}|| \end{aligned}$$

That is, the approximation error committed by $\tilde{f}^{\rm sym}$ is no larger than that of $\tilde{f}$.

8.1 Permutation Invariance

Assume that V_N is permutation invariant, and let $\tilde{V}_N \in {\rm span}\Phi_N$ be an approximation to V_N, and denote the symmetrized approximation by

$$\tilde{V}_N^{\rm perm}(R) \coloneqq \frac{1}{N!} \sum_{\sigma \in S_N} \tilde{V}_N(\sigma R)$$

which we know is at least as accurate as the original approximation $\tilde{V}_N$. We therefore assume from now that $\tilde{V}_N = \tilde{V}_N^{\rm perm}$, i.e. it is already permutation symmetric. Writing

$$\tilde{V}_N = \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} c_{\mathbf{nlm}} \phi_{\mathbf{nlm}}$$

then the linear independence of the $\phi_{\mathbf{n},\mathbf{l},\mathbf{m}}$ and the permutation symmetry $\tilde{V}_N = \tilde{V}_N \circ \sigma$ implies that $c_{\mathbf{n},\mathbf{l},\mathbf{m}} = c_{\sigma\mathbf{n},\sigma\mathbf{l},\sigma\mathbf{m}}$. Therefore, we can write

$$\tilde{V}_N = \sum_{(\mathbf{n},\mathbf{l},\mathbf{m}) \text{ ordered}} c_{\mathbf{nlm}} \sum_{\sigma \in S_N} \phi_{\mathbf{nlm}} \circ \sigma$$

which basically means that we take all ordered triplets $\mathbf{n}, \mathbf{l}, \mathbf{m}$ and sum all of its permutations, which may result in *different* coefficients $c_{\mathbf{nlm}}$.

Definition (Ordered Tuple) We say that a tuple of P vectors $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(P)}) \in \mathbb{Z}^N \times \dots \times \mathbb{Z}^N = (\mathbb{Z}^N)^P$, is ordered if the following vectors of components are ordered lexicographically. That is, we prioritize ordering the first terms, and for vectors with the same first term, we order them using the second term, and so on with the third and fourth, etc.

$$\begin{pmatrix} a_1^{(1)} \\ a_1^{(2)} \\ \vdots \\ a_1^{(P)} \end{pmatrix}, \begin{pmatrix} a_2^{(1)} \\ a_2^{(2)} \\ \vdots \\ a_2^{(P)} \end{pmatrix}, \dots, \begin{pmatrix} a_N^{(1)} \\ a_N^{(2)} \\ \vdots \\ a_N^{(P)} \end{pmatrix}$$

are ordered. For example, let us have the following vectors

$$\begin{pmatrix}1\\2\\5\end{pmatrix}, \begin{pmatrix}1\\2\\-3\end{pmatrix}, \begin{pmatrix}1\\-2\\5\end{pmatrix}, \begin{pmatrix}-3\\2\\5\end{pmatrix}, \begin{pmatrix}2\\0\\5\end{pmatrix}, \begin{pmatrix}5\\2\\-8\end{pmatrix}$$

Then, their ordering would be

$$\begin{pmatrix}-3\\2\\5\end{pmatrix}, \begin{pmatrix}1\\-2\\5\end{pmatrix}, \begin{pmatrix}1\\2\\-3\end{pmatrix}, \begin{pmatrix}1\\2\\5\end{pmatrix}, \begin{pmatrix}2\\0\\5\end{pmatrix}, \begin{pmatrix}5\\2\\-8\end{pmatrix}$$

More importantly, this ordering allows us to represent different permutations of the same group of tuples with a single ordered tuple.

8.2 Point Reflection Invariance

Assume that our approximation $\tilde{V}_N$ of V_N has reflection symmetry, that is, $\tilde{V}_N$ satisfies

$$\tilde{V}_N(R) = \tilde{V}_N(-R)$$

or equivalently, $\tilde{V}_N = \tilde{V}_N \circ J$ where $JR = -R$. Combined rotation and point reflection symmetry is equivalent to O(3), but treating point reflection separately allows us to reduce the number of admissible basis functions. Recall that given the spherical harmonic

$$Y_l^m(\hat{\mathbf{r}}) \coloneqq -P_l^m(\cos\theta)\, e^{mi\phi} \text{ with } \mathbf{r} = r\hat{\boldsymbol{r}} = r(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$$

one of its properties is

$$Y_l^m \circ J = (-1)^l Y_l^m \implies \phi_{nlm} \circ J = (-1)^l \phi_{nlm}$$

and, with $R = (\mathbf{r_1}, \dots, \mathbf{r_N})$, we can therefore write

$$\begin{aligned}
\tilde{V}_N(\mathbf{R}) &= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} c_{\mathbf{nlm}} \frac{1}{2}\big(\phi_{\mathbf{nlm}} + \phi_{\mathbf{nlm}} \circ J\big)(\mathbf{R}) \\
&= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} \frac{1}{2} c_{\mathbf{nlm}} \bigg[\Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big) + \Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big) \circ J\bigg](\mathbf{R}) \\
&= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} \frac{1}{2} c_{\mathbf{nlm}} \bigg[\Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big) + \Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j} \circ J\Big)\bigg](\mathbf{R}) \\
&= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} \frac{1}{2} c_{\mathbf{nlm}} \bigg[\Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big) + \Big(\prod_{j=1}^N (-1)^{l_j} \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big)\bigg](\mathbf{R}) \\
&= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} \frac{1}{2} c_{\mathbf{nlm}} \big[1 + (-1)^{\sum \mathbf{l}}\big]\Big(\prod_{j=1}^N \phi_{\mathbf{n}_j\mathbf{l}_j\mathbf{m}_j}\Big)(\mathbf{R}) \\
&= \sum_{\mathbf{n},\mathbf{l},\mathbf{m}} \frac{1}{2} c_{\mathbf{nlm}} \big(1 + (-1)^{\sum \mathbf{l}}\big)\phi_{\mathbf{nlm}}(\mathbf{R})
\end{aligned}$$

Note that all basis functions $\phi_{\mathbf{nlm}}$ for which $\sum \mathbf{l}$ is odd, vanish under this operation. That is, we only need to retain $(\mathbf{n}, \mathbf{l}, \mathbf{m})$ tuples for which $\sum \mathbf{l}$ is even. The resulting functions with even $\sum \mathbf{l}$ already respect reflection symmetry.

In summary, we have so far shown that we can compute V_N by symmetrised tensor products of form

$$\tilde{V}_N = \sum_{\substack{(\mathbf{n},\mathbf{l},\mathbf{m}) \text{ ordered} \\ \sum \mathbf{l} \text{ even}}} c_{\mathbf{mlm}} \sum_{\sigma \in S_N} \phi_{\mathbf{nlm}} \circ \sigma$$

8.3 Rotation Invariance

Finally, supposing that V_N satisfies the symmetry properties, now including rotation invariance, induces us to rewrite $\tilde{V}_N$ as

$$\tilde{V}_N = \sum_{\substack{(\mathbf{n},\mathbf{l},\mathbf{m}) \text{ ordered} \\ \sum \mathbf{l} \text{ even}}} c_{\mathbf{nlm}} \sum_{\sigma \in S_N} \int_{\mathrm{SO}(3)} \big(\phi_{\mathbf{nlm}} \circ \sigma\big)(Q\hat{\mathbf{R}})\, dQ$$

Recall that

$$\phi_{\mathbf{nlm}}(\mathbf{R}) = \phi_{\mathbf{nlm}}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{\alpha=1}^{N} P_{n_\alpha}(r_\alpha) Y_{l_\alpha}^{m_\alpha}(\hat{\mathbf{r}}_\alpha)$$

we can just focus on integrating over rotations of $Y_{\mathbf{l}}^{\boldsymbol{\mu}}$ rather than $\phi_{\mathbf{nlm}}$, and from (1), recall that for fixed $\mathbf{l} \in \mathbb{N}^N, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}$, and $\hat{\mathbf{R}} = (\hat{\mathbf{r}}_1, \dots, \hat{\mathbf{r}}_N)$, the representation of rotated spherical harmonics in terms of the Wigner D-matrices is

$$Y_{\mathbf{l}}^{\mathbf{m}}(Q\hat{\mathbf{R}}) = \sum_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, Y_{\mathbf{l}}^{\boldsymbol{\mu}}(\hat{\mathbf{R}}), \text{ where } D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) = \prod_{\alpha=1}^{N} D^{l_\alpha}_{\mu_\alpha m_\alpha}(Q)$$

We wish to make the $\phi_{\mathbf{nlm}}$'s, and therefore the $Y_{\mathbf{l}}^{\mathbf{m}}$'s symmetric. In general, if we have a function f, we can apply the symmetric operator

$$\mathcal{S} : f \mapsto \bar{f} = \int_{\mathrm{O}(3)} f \circ Q \, dQ$$

resulting in $\bar{f} = \mathcal{S}(f)$ such that $\bar{f}(Q\mathbf{r}) = \bar{f}(\mathbf{r})$ for all $\mathbf{r} \in \mathbb{R}^3$. We apply the same operator to the function $Y_{\mathbf{l}}^{\mathbf{m}}$:

$$\begin{aligned}
\mathcal{S}(Y_{\mathbf{l}}^{\mathbf{m}}) = \int_{\mathrm{O}(3)} Y_{\mathbf{l}}^{\mathbf{m}}(Q\hat{\mathbf{R}})\, dQ &= \int_{\mathrm{O}(3)} \Big(\sum_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) Y_{\mathbf{l}}^{\boldsymbol{\mu}}(\hat{\mathbf{R}}) \Big)\, dQ \\
&= \sum_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} \Big(\underbrace{\int_{\mathrm{O}(3)} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, dQ}_{\bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}} \Big) Y_{\mathbf{l}}^{\boldsymbol{\mu}}(\hat{\mathbf{R}}) \\
&= \sum_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} \bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}} Y_{\mathbf{l}}^{\boldsymbol{\mu}}(\hat{\mathbf{R}})
\end{aligned}$$

where the integrated coefficients $\bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}$, also called coupling coefficients, can be computed efficiently using the recursive formula (A.11) in the ACE paper with Lemma 3.

Proposition (ACE Lemma 3) For $N \geq 2$ and $\mathbf{l} \in \mathbb{N}^N$, let

$$\mathcal{L}_{\mathbf{l}} \coloneqq \Big\{ (L_1, L_2, \dots, L_N) \in \mathbb{N}^N \,\Big|\, L_1 = l_1 \text{ and } |L_{i-1} - l_i| \leq L_i \leq L_{i-1} + l_i,\ 2 \leq i \leq N \Big\}$$

where it is assumed that $l_0 = 0$. Let $\boldsymbol{\mu}, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}$. Then,

$$D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) = \sum_{\mathbf{L} \in \mathcal{L}_N} [\mathcal{C}_{\mathbf{l}}]_{\mathbf{m},\mathbf{L}} [\mathcal{C}_{\mathbf{l}}]_{\boldsymbol{\mu},\mathbf{L}} D^{L_N}_{\tilde{M}_N M_N}(Q)$$

with

$$\begin{aligned}
M_N &\coloneqq \sum_{j=1}^{N} m_i \\
\tilde{M}_N &= \sum_{j=1}^{N} \mu_i \\
[\mathcal{C}_{\mathbf{l}}] &\coloneqq \mathcal{C}^{L_2 M_2}_{l_1 m_1 l_2 m_2} \mathcal{C}^{L_3 M_3}_{L_2 M_2 l_3 m_3} \dots \mathcal{C}^{L_N M_N}_{L_{N-1} M_{N-1} l_N m_N}
\end{aligned}$$

With this formula, we can reduce the integral of the vector-indexed $D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)$ into the sums of integrals of simpler D-matrices.

$$\bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q) = \int_{\mathrm{O}(3)} D^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}(Q)\, dQ = \sum_{\mathbf{L}\in\mathcal{L}_N} [\mathcal{C}_{\mathbf{l}}]_{\mathbf{m},\mathbf{L}} [\mathcal{C}_{\mathbf{l}}]_{\boldsymbol{\mu},\mathbf{L}} \left(\int_{\mathrm{O}(3)} D^{L_N}_{\tilde{M}_N M_N}(Q)\, dQ \right)$$

where we can define the block matrix

$$D^{\mathbf{l}}(Q) \coloneqq \bigoplus_{\mathbf{L}\in\mathcal{L}_{\mathbf{l}}} D^{L_N}(Q)$$

It follows that

$$\bar{D}^{L_N}_{\tilde{M}_N M_N}(Q) = \int_{\mathrm{O}(3)} D^{L_N}_{\tilde{M}_N M_N}(Q)\, dQ = \begin{cases} 1 & \text{if } L_N = \tilde{M}_N = M_N = 0 \\ 0 & \text{if else} \end{cases}$$

since by symmetry of $\bar{D}^{L_N}_{\tilde{M}_N M_N}$, the integrals end up vanishing to 0 if the indices are not all 0. Therefore, we just look at the components $\bar{D}^{L_N}_{\tilde{M}_N M_N}$ that equals to 1 multiply it by the generalized Clebsch-Gordan coefficients $\mathcal{C}_{\mathbf{l}}$, and sum over $\mathcal{L}_N$ to get each coefficient $\bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}$. Remember that these values $\bar{D}^{L_N}_{\tilde{M}_N M_N}$ represent the components of the symmetrised D-matrix $\bar{D}^{L_N}$ which is shown to be sparse (many zeros).

As an example, for $\mathbf{l} = (1,1,2), \mathcal{L}_{\mathbf{l}} = \{(1,1),(2,0),(2,1),(2,2),(2,3)\}$. Then, we have the 5 matrices $D^{L_N}(Q)$:

$$D^1(Q),\ D^0(Q),\ D^1(Q),\ D^2(Q),\ D^3(Q)$$

with each $D^i(Q)$ being size $(2i+1)\times(2i+1)$. Letting $D^{\mathbf{l}}(Q) \coloneqq D^1(Q) \oplus D^0(Q) \oplus D^1(Q) \oplus D^2(Q) \oplus D^3(Q)$, we have the block matrix

$$D^{\mathbf{l}}(Q) = \begin{pmatrix} D^1(Q) & 0 & 0 & 0 & 0 \\ 0 & D^0(Q) & 0 & 0 & 0 \\ 0 & 0 & D^1(Q) & 0 & 0 \\ 0 & 0 & 0 & D^2(Q) & 0 \\ 0 & 0 & 0 & 0 & D^3(Q) \end{pmatrix}$$

where the 0s are zero matrices of matching size. Moreover, symmetrising the elements gives

$$\int_{\mathrm{SO}(3)} D^{\mathbf{l}}(Q)\, dQ = \begin{pmatrix} 0_{3,3} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0_{3,3} & 0 & 0 \\ 0 & 0 & 0 & 0_{5,5} & 0 \\ 0 & 0 & 0 & 0 & 0_{7,7} \end{pmatrix}$$

where the size of the matrices are made specific on the diagonal. Multiplying by the Clebsch coefficients results in the matrix $\bar{D}^{\mathbf{l}}(Q)$ of relevant coefficients. We can see that in this case, $\mathrm{rank}(\bar{D}^{(1,1,2)}) = \tilde{n}_{\mathbf{l}} = 1$.

Stepping back from calculating the coefficients, performing this symmetric operator $\mathcal{S}$ on all the $\{Y^{\mathbf{m}}_{\mathbf{l}}\}$ gives the symmetrised set of functions $\{b_{\mathbf{lm}}\}$, where

$$b_{\mathbf{lm}} = \mathcal{S}(Y^{\mathbf{m}}_{\mathbf{l}})$$

In matrix form, for fixed $\mathbf{l}$, the symmetrised Wigner D-matrix $\bar{D}^{\mathbf{l}}$ with components $\bar{D}^{\mathbf{l}}_{\boldsymbol{\mu}\mathbf{m}}$ is a $|\mathcal{M}_{\mathbf{l}}| \times |\mathcal{M}_{\mathbf{l}}|$ linear operator mapping the $|\mathcal{M}_{\mathbf{l}}$-dimensional vector of regular spherical harmonics $\left(Y^{\boldsymbol{\mu}}_{\mathbf{l}}\right)_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}}$ to the $|\mathcal{M}_{\mathbf{l}}|$-dimensional vector of symmetrised harmonics $\left(b_{\mathbf{lm}}\right)_{\mathbf{m}\in\mathcal{M}_{\mathbf{l}}}$.

$$\underbrace{\begin{pmatrix} & \bar{D}^{\mathbf{l}} & \end{pmatrix}}_{|\mathcal{M}_{\mathbf{l}}|\times|\mathcal{M}_{\mathbf{l}}|} \underbrace{\begin{pmatrix} | \\ Y^{\boldsymbol{\mu}}_{\mathbf{l}} \\ | \end{pmatrix}_{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}}}_{|\mathcal{M}_{\mathbf{l}}|\times 1} = \underbrace{\begin{pmatrix} | \\ b_{\mathbf{lm}} \\ | \end{pmatrix}_{\mathbf{m}\in\mathcal{M}_{\mathbf{l}}}}_{|\mathcal{M}_{\mathbf{l}}|\times 1}$$

This set of components $\{b_{\mathbf{lm}}\}$ spans the set of symmetric functions, but it is not necessarily linearly independent, so we must transform it into an orthonormal basis. The following lemma helps us do that. It colloquially states that rather than using the symmetrised Wigner D-matrix $\bar{D}^{\mathbf{l}}$, we can construct another linear map $\tilde{\mathcal{U}}^{\mathbf{l}} : \mathbb{R}^{|\mathcal{M}_{\mathbf{l}}|} \longrightarrow \mathbf{R}^{\tilde{n}_{\mathbf{l}}}$ (with specific properties stated below) that produces a $\tilde{n}_{\mathbf{l}}$-dimensional vector ($\tilde{n}_{\mathbf{l}} \leq |\mathcal{M}_{\mathbf{l}}|$) of orthonormal functions $\{b_{\mathbf{l}i} \,|\, i = 1, \ldots, \tilde{n}_{\mathbf{l}}\}$ that form a basis of $\text{span}\{b_{\mathbf{lm}} \,|\, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}\}$.

$$\underbrace{\begin{pmatrix} & \tilde{\mathcal{U}}^{\mathbf{l}} & \end{pmatrix}}_{\tilde{n}_{\mathbf{l}} \times |\mathcal{M}_{\mathbf{l}}|} \underbrace{\begin{pmatrix} | \\ Y_{\mathbf{l}}^{\boldsymbol{\mu}} \\ | \end{pmatrix}_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}}}_{|\mathcal{M}_{\mathbf{l}}| \times 1} = \underbrace{\begin{pmatrix} | \\ b_{\mathbf{l}i} \\ | \end{pmatrix}_{i=1,\ldots,\tilde{n}_{\mathbf{l}}}}_{\tilde{n}_{\mathbf{l}} \times 1}$$

We find $\tilde{n}_{\mathbf{l}}$ computationally.

Proposition (ACE Lemma 2) Suppose that we can find $\tilde{\mathcal{U}}^{\mathbf{l}}$ a matrix with orthonormal rows $\tilde{\mathcal{U}}^{\mathbf{l}}_i$ for $i = 1, \ldots \tilde{n}_{\mathbf{l}}$ (with $\tilde{n}_{\mathbf{l}} = \text{rank}\bar{D}^{\mathbf{l}}$) each of dimension $|\mathcal{M}_{\mathbf{l}}|$. That is, we can interpret $\tilde{\mathcal{U}}^{\mathbf{l}}_i = (\tilde{\mathcal{U}}^{\mathbf{l}}_{i\boldsymbol{\mu}})_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} \in \mathbb{R}^{|\mathcal{M}_{\mathbf{l}}|}$ as a $|\mathcal{M}_{\mathbf{l}}|$-dimensional vector indexed by μ, and

$$\tilde{\mathcal{U}}^{\mathbf{l}} = \begin{pmatrix} — & \tilde{\mathcal{U}}^{\mathbf{l}}_1 & — \\ \vdots & \vdots & \vdots \\ — & \tilde{\mathcal{U}}^{\mathbf{l}}_{\tilde{n}_{\mathbf{l}}} & — \end{pmatrix} \in \mathbb{R}^{\tilde{n}_{\mathbf{l}} \times |\mathcal{M}_{\mathbf{l}}|}$$

and therefore $\tilde{\mathcal{U}}^{\mathbf{l}}$ is a linear map from $\mathbb{R}^{|\mathcal{M}_{\mathbf{l}}|} \longrightarrow \mathbb{R}^{\tilde{n}_{\mathbf{l}}}$. Suppose further that $\text{Im}(\tilde{\mathcal{U}}^{\mathbf{l}}) = \text{Im}(\bar{D}^{\mathbf{l}}$. Then, $\tilde{\mathcal{U}}^{\mathbf{l}}$ generates an orthonormal basis. That is, the functions

$$b_{\mathbf{l}i}(\hat{\mathbf{R}}) \coloneqq \sum_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}} \tilde{\mathcal{U}}^{\mathbf{l}}_{\boldsymbol{\mu}i} Y_{\mathbf{l}}^{\boldsymbol{\mu}}(\hat{\mathbf{R}}), \qquad i = 1, 2, \ldots, \tilde{n}_{\mathbf{l}}$$

form an orthonormal basis of $\text{span}\{b_{\mathbf{lm}} \,|\, \mathbf{m} \in \mathcal{M}_{\mathbf{l}}\}$. Matrix-wise, $\tilde{\mathcal{U}}^{\mathbf{l}}$ maps the unsymmetrised harmonics to the vector contains the orthonormal basis functions $b_{\mathbf{li}}$.

$$\underbrace{\begin{pmatrix} & \tilde{\mathcal{U}}^{\mathbf{l}} & \end{pmatrix}}_{\tilde{n}_{\mathbf{l}} \times |\mathcal{M}_{\mathbf{l}}|} \underbrace{\begin{pmatrix} | \\ Y_{\mathbf{l}}^{\boldsymbol{\mu}} \\ | \end{pmatrix}_{\boldsymbol{\mu} \in \mathcal{M}_{\mathbf{l}}}}_{|\mathcal{M}_{\mathbf{l}}| \times 1} = \underbrace{\begin{pmatrix} | \\ b_{\mathbf{l}i} \\ | \end{pmatrix}_{i=1,\ldots,\tilde{n}_{\mathbf{l}}}}_{\tilde{n}_{\mathbf{l}} \times 1}$$

9 Efficient Evaluation

Given an atomic neighborhood $R = \{\mathbf{r}_j\}_{j=1}^J$, we can define the **atomic density** of the particles to be

$$\rho_R(\mathbf{r}) \coloneqq \sum_{j=1}^{J} \delta(\mathbf{r} - \mathbf{r}_j)$$

where δ is the Dirac delta function (in a way, we can interpret it as a Gaussian mixture model, but with delta functions). Then, we choose a one-particle basis

$$\phi_{nlm}(\mathbf{r}) = R_n(r) Y_l^m(\hat{\mathbf{r}})$$

and project ρ onto that basis

$$\begin{aligned} A_{nlm}(R) &\coloneqq \langle \phi_{nlm}, \rho_R \rangle \\ &= \int \phi_{nlm}(r) \sum_{j=1}^{J} \delta(\mathbf{r} - \mathbf{r}_j)\, dr \\ &= \sum_{j=1}^{J} \underbrace{\int \phi_{nlm} \delta(\mathbf{r} - \mathbf{r}_j)\, dr}_{\phi_{nlm}(\mathbf{r}_j)} \\ &= \sum_{j=1}^{J} \phi_{nlm}(\mathbf{r}_j) \end{aligned}$$

It is not clear to see the motivation behind taking the projection of ρ onto the basis, so just take it for granted here. Next, we form the N-correlations of the density $\rho_R^{\otimes N}$ and project them onto the multi-particle (tensor product) basis

$$A_{\mathbf{nlm}}(R) \coloneqq \left\langle \bigotimes_{\alpha=1}^{N} \phi_{n_\alpha l_\alpha m_\alpha},\ \rho^{\otimes N} \right\rangle = \prod_{\alpha=1}^{N} A_{n_\alpha l_\alpha m_\alpha}(R)$$

Finally, we symmetrize the N-correlations by integrating over the O(3) Haar measure,

$$B_{\mathbf{nlm}} \propto \int_{\mathrm{O}(3)} A_{\mathbf{nlm}} \circ Q\, dQ$$

Because of properties of the spherical harmonics one can write this using the Wigner D-matrices as $B = \mathcal{U}A$. If one symmetrised all possible N-correlations then this creates a spanning set, but one can easily reduce this to an actual basis. This construction then yields a basis of the space of symmetric polynomials, which we index with $B_{\mathbf{nlm}}$. After a few steps that gets rid of linear independence, we index it as $B_{\mathbf{nl}i}$ to denote that i it is a basis. Ultimately, we have

$$B_{\mathbf{nl}i}(\mathbf{R}) \coloneqq \sum_{\mathbf{m} \in \mathcal{M}_\mathbf{l}} \mathcal{U}^{\mathbf{nl}}_{\mathbf{m}i} A_{\mathbf{nlm}}(\mathbf{R})$$

where A is the vector of the correlations $\big(A_{\mathbf{nlm}}\big)_{\mathbf{m} \in \mathcal{M}_\mathbf{l}}$, B is the vector of the new basis $\big(B_{\mathbf{nl}i}\big)_{i=1,\dots,n_{\mathbf{nl}}}$, and $\mathcal{U} \in \mathbb{R}^{n_{\mathbf{nl}} \times |\mathcal{M}_\mathbf{l}|}$ is a sparse matrix of the coupling coefficients. Apparently this construction avoids the $N!$ cost for symmetrising the basis, making it much more computationally efficient.

10 Finnis-Sinclair Model

Now recalling the potential energy surface (PES) of a configuration of particles, we can decompose the PES as:

$$E\big(\{r_i\}\big) = \sum_{i<j} V(\mathbf{r}_{ij}) + \sum_i F\bigg(\sum_{j \neq i} \rho(\mathbf{r}_{ij}) \bigg)$$

with maximal order of interaction truncation to be $\mathcal{N} = 1$ and no external force ($V_0 = 0$). The first summation is simply the sums of the lower-order potential functions, but the second summation of the forces of the electronic density ρ (not related to the atomic density in the previous section) can be regarded as intrinsically as part of the model. Furthermore, since these potential functions exhibit rotation, reflection, permutation symmetry, the potentials V and the densities ρ are reduced to being dependent only on the distances r_{ij}.

$$E\big(\{r_i\}\big) = \sum_{i<j} V(r_{ij}) + \sum_i F\bigg(\sum_{j \neq i} \rho(r_{ij}) \bigg)$$

1. The ACE paper states that the potential functions can be approximated to arbitrarily accuracy by the countably infinite basis P_k of radial polynomials. We approximate it with a finite K number of basis functions.
$$V(\mathbf{r}) \approx \sum_{k=1}^{K} a_k P_k(|\mathbf{r}|)$$

2. It is assumed that the smooth density ρ can be decomposed into a finite K' subbasis of radial polynomials.
$$\rho(\mathbf{r}) \approx \sum_{k=1}^{K'} a'_k P'_k(|\mathbf{r}|)$$

3. The **Finnis-Sinclair model** assumes that $F(x) = -\sqrt{x}$.

Note that this model consists of only pairwise interactions, i.e. it is the sum of two-body interactions $V(r_{ij})$ and $\rho(r_{ij})$. We introduce a new, more general model that consists of higher-order body interactions that isn't limited to two:
$$E\big(\{r_i\}\big) = \sum_i \hat{V}\big(\{r_{ij}\}\big) + \sum_i F\big[\hat{\rho}\big(\{r_{ij}\}\big)\big]$$
which would be equivalent to the previous model if
$$\underbrace{\hat{V}\big(\{r_{ij}\}_{j\neq i}\big)}_{\text{multi-body function}} = \sum_{i\neq j} \frac{1}{2} V\big(r_{ij}\big) \qquad \text{(i.e. } V \text{ only has pairwise interactions)}$$
$$\hat{\rho}\big(\{r_{ij}\}_{j\neq i}\big) = \sum_{j\neq i} \frac{1}{2} \rho(|r_{ij}|)$$

Recall that for the single-body $V(\mathbf{r}_{ij})$, we could decompose it to $V(r_{ij})$ by imposing symmetries. However, for the more general multi-body model $\hat{V}\big(\{\mathbf{r}_{ij}\}\big)$, it is *not* just a function of scalar distances $V\big(\{r_{ij}\}\big)$ because we need to account for angular information. Reducing the vector coordinates scalar data will wash away the angular information. Using the symmetrised ACE basis functions $\mathcal{B}$ constructed in the ACE paper and approximating it with a finite subbasis, we have
$$\hat{V}\big(\{\mathbf{r}_{ij}\}\big) \approx \sum_{k=1}^{K} a_k \mathcal{B}_k\big(\{\mathbf{r}_{ij}\}\big)$$
$$\hat{\rho}\big(\{\mathbf{r}_{ij}\}\big) \approx \sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)$$
and therefore we can approximate $E\big(\{\mathbf{r}_i\}\big) \approx \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big)$ where, by substituting the approximations above, we get
$$\hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big) = \sum_i \sum_{k=1}^{K} a_k \mathcal{B}_k\big(\{\mathbf{r}_{ij}\}\big) + \sum_i F\bigg(\sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)\bigg)$$
Notice that $\mathbf{a}, \mathbf{a}'$ are K, K'-dimensional vectors representing the coefficients of the linear combinations of $\hat{V}$ and $\hat{p}$. Again, like the single-body model, if we choose basis $\mathcal{B}$ considering only pairwise interactions, this model would just collapse to the pairwise model. Implemented in `evaluate`.

10.1 Differentiation

We would like to implement this function $\hat{E}\big(\{r_i\}\big) = \hat{E}(R)$ in Julia. We first compute the gradient.

1. Finding $\nabla_R \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big)$ is easy.

2. Finding $\nabla_a \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big)$ requires us to look at only the first summation, since the second one vanishes. It is a vector of K elements:

$$\begin{aligned}\nabla_a \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big) &= \Big(\sum_i \mathcal{B}_k\big(\{\mathbf{r}_{ij}\}\big) \Big)_{k=1}^{K} \\ &= \sum_i \underbrace{\Big(\mathcal{B}_k\big(\{\mathbf{r}_{ij}\}\big) \Big)_{k=1}^{K}}_{\texttt{[b.val for b in B1]}}\end{aligned}$$

Implemented in `grad1` in the `ACE.param_evaluate_d`.

3. Finding $\nabla_{a'} \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big)$ is slightly harder, but since $F(x) = -\sqrt{x}$, with chain rule we see that

$$\partial_{a'_k} \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big) = \sum_i \frac{1}{2} \frac{1}{\sqrt{\sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)}} \cdot \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)$$

and so, in gradient form, we have a K'-dimensional vector:

$$\begin{aligned}\nabla_{a'} \hat{E}\big(\{\mathbf{r}_i\}, \mathbf{a}, \mathbf{a}'\big) &= \left(\sum_i \frac{\mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)}{2\Big(\sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big) \Big)^{1/2}} \right)_{k=1}^{K'} \\ &= \sum_i \left(\frac{\mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big)}{2\Big(\sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big) \Big)^{1/2}} \right)_{k=1}^{K'} \\ &= \sum_i \underbrace{\frac{1}{2\Big(\sum_{k=1}^{K'} a'_k \mathcal{B}'_k\big(\{\mathbf{r}_{ij}\}\big) \Big)^{1/2}}}_{\substack{\texttt{m.transform\_d(sum(c*b for (c,b)} \\ \texttt{in zip(m.c2,B2)).val)}}} \underbrace{\Big(\mathcal{B}'_k\big(\mathbf{r}_{ij}\big) \Big)_{k=1}^{K'}}_{\texttt{[b.val for b in B2]}}\end{aligned}$$

Implemented in `grad2` in the `ACE.param_evaluate_d`.